

No farewell to arms?

If North Korea is to bid farewell to nuclear arms, negotiators will have to convince the country that its neighbours do not themselves want to become nuclear powers. Recent activities in Japan and South Korea will make this harder.

It's easy for a totalitarian state to produce propaganda. There is no independent media, after all, to correct any lies. North Korea's nuclear weapons programme is shrouded in half-truths — mainly claims that its existence is necessary because neighbouring countries have similar ambitions. These allegations are probably false, but they would be easier to ignore if Japan and South Korea stopped fuelling North Korea's misinformation machine.

Earlier this month, for example, South Korea admitted that its scientists conducted experiments in 2000 on enriching uranium, and around two decades earlier on extracting plutonium from nuclear waste. By not reporting either project to the International Atomic Energy Agency (IAEA), the South violated the Treaty on the Non-Proliferation of Nuclear Weapons. The fact that this didn't come out until it was discovered by IAEA inspectors makes matters worse.

North Korea is stalling on attempts to resume talks over its weapons plans, and quickly used the revelation as proof that it needs a nuclear capability. The country is so adept at diplomatic smoke and mirrors that nuclear analysts are unsure of the size of its nuclear arsenal, or indeed if it has one at all. But it is clear that North Korea has an active nuclear programme, and the South Korean revelations have dimmed hopes that talks aimed at ending that work will resume in the coming months.

Sadly, this is not the first own-goal scored by North Korea's neighbours. In January 2003, Japan revealed that 200 kilograms of used plutonium was missing from its nuclear power complexes. Japan is also building a massive reprocessing plant for spent fuel, which could be used to produce plutonium. This, as North Korea

has repeatedly pointed out, is a fuel for nuclear weapons as well as for nuclear reactors.

In truth, there is no evidence that either South Korea or Japan has a centralized weapons development programme. The hole in Japan's nuclear accounts was due to shoddy monitoring, not under-hand weapons plans, the IAEA said following an inspection a month after the revelation. The agency gave the country another vote of confidence this month by halving the size of its inspection team there.

North Korea's more serious allegations about South Korea will probably also prove baseless when an IAEA team reports on the clandestine experiments next month. But the South cannot be completely innocent. Somewhere down the line, someone decided to push the research and scientists played along. Claims that the experiments were conducted in the name of basic research seem hard to believe given their obvious potential weapons applications. Many industrialized countries have the resources and expertise to conduct such experiments. Maintaining such a capability does not breach the non-proliferation treaty; doing the experiments clearly does.

On the Korean peninsula, such breaches give North Korea another reason to insist on its right to build nuclear weapons. Mismanaging nuclear resources, even civilian ones, has the same effect. That can seem a harsh lesson to have to learn when dealing with a country as manipulative and dishonest as North Korea. But if the region is to avoid a nuclear arms race, the scientists and rulers of countries that have sworn not to build nuclear weapons must stick firmly to the IAEA's rules. ■

Time to look to the future

Germany is understandably cautious about embryo research, but the country would benefit from joint European projects.

After much debate and psychological anguish, the film *Der Untergang* (The Downfall) was released last week in Germany. It tells the story of the last 12 days of Hitler's life, before his suicide in April 1945. In portraying Hitler as a human being rather than a one-dimensional caricature, the film breaks a 60-year taboo, and is a measure of the evolution in Germany's attitudes to its past.

Other attitudes shaped by the Nazi period have evolved even more slowly. Last week saw the release of a report on human cloning by the influential German National Ethics Council. Its cautious tone illustrates how slow has been the evolution of attitudes towards the sanctity of life, which have been so deeply influenced by the Nazi abuse of genetics. In no other Western country is the spectrum of attitudes towards cloning so narrow, and so skewed towards conservatism.

No taboo has been broken in the report. The 25 council members, who include theologians, physicians and scientists, unanimously called for a worldwide ban on reproductive cloning. But they declined to take a vote on cloning for research purposes, going only as far as to say whether, as individuals, they supported a general ban, a temporary ban or limited approval. The German government,

which established the council three years ago, was quick to jump on the general recommendation of a temporary continuation of the existing ban.

The same cautious approach was responsible for Germany's long delay in allowing any form of genetic engineering, and its late entry to the international Human Genome Project. Since then there has been pressure to catch up. It is historically understandable that Germany holds back on any work involving human embryos, but, given the increasingly clear potential of therapeutic cloning and work on human embryonic stem cells, it is important for small steps to be taken.

One such step would be for Germany to relax its rules on the use of cultured human embryonic stem-cell lines. At present, German scientists may not work on cell lines created after 2002, when the rules were fixed. However, older cell lines are of limited use in research. The European Union's Sixth Framework Programme funds research on new cultured lines, and there are German researchers who want to take part in such projects. The country will one day want to enter this field, as it did in genomics. When it does, expertise needs to be on hand. ■





Science head
Alan Bement set to
lead National Science
Foundation
p388



Ethical dilemma
Permission
sought for face
transplants
p389



Undercover
Computers help
to reconstruct
hidden Picasso
p391



Wildlife protest
Japan's rare dugongs
threatened by plan
for army base
p392

European space scientists brim with ideas for going it alone

Quirin Schiermeier, Munich

Europe's top space researchers are calling on the European Space Agency (ESA) to develop a bolder vision for space research and exploration.

At a meeting held in Paris last week, researchers said that they wanted the agency to develop a scientific profile that would be more distinct from that of NASA, the agency's US counterpart. Many urged ESA to embrace space-based experiments in particle physics and the search for extraterrestrial life, in addition to its traditional priorities of astronomy and planetary exploration.

Some 400 researchers, including most of Europe's leading space scientists, attended the Cosmic Vision 2015–2025 meeting. Such meetings are held every ten years to collect ideas for ESA's science plan, the next of which is scheduled to be published next year.

Some researchers at the meeting called on ESA to rethink its existing mix of collaboration and competition with NASA. Their doubts about the US agency's leadership have been fuelled by uncertainty about the International Space Station, confusion over NASA's plans to return to the Moon and to Mars, and the agency's propensity to hog the limelight for results obtained from joint projects.

European planetary scientists are especially keen to break free from NASA's influence. "We shouldn't be deterred by what NASA plans to do," says Wolfgang Baumjohann, director of the Space Research Institute in Graz, Austria, who chairs ESA's working group on Solar System science. "We want to go outward ourselves and explore the Solar System beyond Jupiter and its satellites — and we also want to land on moons and planets, and return samples."

Researchers are particularly keen to explore Europa, one of Jupiter's moons, which is thought to harbour a vast lake beneath a thick crust of ice. NASA has already proposed a similar mission, the Jupiter Icy Moons Orbiter. This revolutionary spacecraft — part of the US agency's Project Prometheus — would be powered by nuclear fission. But doubts have been expressed over the mission's feasibility (see *Nature* 431, 113; 2004). European scientists are hoping that



Higher profile: European scientists want to mount a solo mission to Europa.

ESA will back a cheaper and less technically challenging Europa mission, fuelled by conventional power sources.

European physicists, meanwhile, would like to build space-based experiments, says Bernard Schutz, a director at the Max Planck Institute for Gravitational Physics near Potsdam, Germany, and chair of ESA's fundamental physics advisory group.

For example, researchers would like to use an array of spacecraft, communicating with each other by laser, to detect gravitational waves — faint disturbances in the curvature of space-time predicted by Albert Einstein. By monitoring the distances between the spacecraft, scientists hope to detect the mysterious ripples, which are thought to emanate from major cosmological events such as collisions between black holes, as well as from the Big Bang. In addition, space-based quantum experiments are being designed to measure phenomena such as possible slow variations in fundamental physical constants over time.

"These things need to be explored in space," says Schutz. "We just can't make lab experiments last long enough to gain any clarity."

Competition between the United States and Europe is also driving the search for

habitable planets outside the Solar System. European scientists hope that advanced observation tools will allow them to detect biomarkers in the atmosphere of planets around distant stars, says Catherine Turon, an astronomer at the Paris Observatory at Meudon.

At the meeting, researchers presented 150 different ideas that will now be considered in more detail by ESA advisory committees for possible inclusion in next year's plan. The meeting drew twice as many proposals as the previous one in 1994.

"I am delighted about the number and quality of the ideas brought forward," says Giovanni Bignami, chairman of ESA's space science advisory committee. "But it is obvious that we still have too much vision for even the most rosy predictions of budget increases."

From its total annual budget of €2.7 billion (US\$3.3 billion), ESA currently spends €370 million per year on science — an order of magnitude less than NASA, although the European agency's money goes a long way because most researchers' salaries are paid by their home nations. In the next six months, ESA's advisory groups will analyse and revise the proposals to a reasonable size, so that they will fit in with the agency's limited resources.

Bush backs Bement to head science agency

Geoff Brumfiel, Washington

Arden Bement, a metallurgist with a strong track record in industry, government and the academic world, has been nominated by the US president, George W. Bush, to serve as the National Science Foundation's twelfth director.

Bement is currently director of the National Institute of Standards and Technology (NIST), but has been doing double duty as acting director of the US science agency since marine biologist Rita Colwell stepped down in February (see *Nature* 427, 665; 2004).

The 15 September nomination was made three days before a legal deadline for nominating a permanent director. It awaits confirmation by the Senate.

Bush put Bement's name forward after months of fruitless searching for a permanent replacement for Colwell, observers say. The search was hampered by uncertainty ahead of November's presidential election, as candidates did not know which president they would be serving — or even whether their nomination would stand if Bush lost.

The National Science Foundation is the main US funding agency for university research in disciplines other than biomedical science. Its directors are often chosen for



US President Bush greets Arden Bement, who supporters hope will 'get things done' at the National Science Foundation.

their administrative experience: Colwell, an accomplished scientist with big ideas for the agency's expansion, was an exception.

Supporters of the agency are upbeat about 72-year-old Bement's appointment. "He's a team player who knows how to get things done within the government," one science lobbyist said.

Before his four-year stint as director of NIST, Bement was head of the nuclear engineering department at Purdue University, Indiana, and had previously run research and development at TRW, a major military contractor. He is no stranger to Washington, having led the materials science programme

at the Defense Advanced Research Projects Agency and served briefly, in 1979, as head of research and development at the Pentagon.

"I'm quite eager to accept this responsibility," Bement says. As director of the agency, he adds, he will work to bring more women and members of minority groups to the sciences. In addition, he hopes to increase the success rate for grant applications.

He also hopes to take up the appointment whoever wins the election. "Historically, the foundation has responded to the needs of science, regardless of the administration," he says. "I would want to preserve this tradition."

No serious opposition to Bement is likely in the Senate, but when the nomination will be confirmed remains uncertain. Congress is on a shortened schedule and is struggling to deal with more pressing nominations, and Democrats may seek to block new appointments until after the election, in case Kerry wins and wants to make his own choices.

"It is still possible that we could do the confirmation this year," says Gayle Osterberg, a Senate spokeswoman. But others are less sanguine. "I don't know if we would confirm a dog catcher right now," opines another Senate staffer. ■

Hurricane Ivan highlights future risk for New Orleans

Tony Reichhardt, Washington

Tropical storms and hurricanes have surged through much of the Caribbean this month, leaving hundreds dead in flooded Haiti, devastating several other islands and causing havoc along parts of the US coast.

But scientists say that another disaster was only narrowly avoided — Hurricane Ivan missed the deeply vulnerable city of New Orleans by a tiny margin. In the face of future such storms, they are calling for action to restore the area's wetlands, to act as a barrier against flooding.

The Louisiana city was directly in Ivan's path on 14 September, when the storm was classed as 'Category 5' with winds upward of 240 kilometres per hour. Almost a million people were evacuated. But Ivan's centre came ashore more than 160 kilometres east of New Orleans on 16 September, and no major flooding was reported in the city.

Scientists warn that a storm is bound to hit New Orleans at some point in the near future, however. Simulations run by the Center for the Study of Public Health Impacts of Hurricanes at Louisiana State University in



Getting the wind up: Ivan battered outer New Orleans, but experts fear a direct hit in future.

Baton Rouge show that direct hits by even Category 3 storms would have disastrous consequences. Storm surges could flood the city, which lies in a basin mostly below sea level, to a depth of six metres. This would take weeks to drain. "You're talking about total destruction," says the centre's director Ivor van Heerden, claiming that deaths could be in the tens of thousands.

Greg Stone, director of the Coastal Studies Institute at Louisiana State

University, thinks the city's levees would hold after an assault from a Category 3 hurricane — but not from a Category 4 or 5 storm. Both scientists agree that the mayor was correct in urging the city's 1.3 million residents to flee.

Restoring barrier reefs and marshlands in the delta between the city and the Gulf of Mexico is the best way to mitigate the threat, the researchers say. The wetlands have been steadily eroded since the 1930s owing to extensive embankments built to control flooding of the Mississippi river. These prevent water and sediment from reaching the delta. "We are literally starving our wetlands to death," says van Heerden.

A 30-year restoration plan, called Coast 2050, was published in 1998. But arguments over who should foot the \$14-billion bill have stalled the plan in Congress ever since. Stone and other scientists were planning to brief Louisiana's congressional delegation on coastal restoration this week.

The hurricane season, which is currently at the peak of a 30–40 year cycle, is expected to continue for a few more months. ■

Surgeons seek go-ahead to perform first face transplant

Erika Check, Washington

A surgical team pushing for permission to do the world's first face transplant said last week that the potential benefits of the procedure outweigh the social and ethical problems that it raises.

Ethical bodies in France and Britain have already denied permission for surgeons in those countries to perform face transplants, which would probably be used to treat people whose features had been severely disfigured in fires or other accidents.

But now a team of surgeons, psychologists and sociologists at the University of Louisville in Kentucky and Utrecht University in the Netherlands says that it can address the concerns raised by ethical boards elsewhere. In an article in the *American Journal of Bioethics* on 17 September, the team argues that science is far enough advanced to diminish the risks posed by the operation, pointing to advances in immunosuppressive medication and in plastic surgery on complicated organs such as hands.

The team has already submitted a face-transplant proposal to an ethics review body in the Netherlands, and it plans to submit one soon to a review board in the United States. John Barker of the University of Louisville, who leads the team, says he hopes to hear whether the Dutch proposal has been accepted by the end of this week.

In the article, the team admits that there are many uncertainties about face transplants. For instance, the face is intimately linked with a person's identity, and it is difficult to know how a face-transplant recipient will adjust to his or her new appearance. And the family of the patient who donates the face might take exception to the procedure. But proponents say that there is a limit to what can be done to address these issues, which cannot be resolved through further experiments on animals or lab tests. "There arrives a point in time when the procedure should simply be done," the team writes. "We submit that that time is now."

However, ethicists and specialists asked to review the proposal have said that it is premature. In responses published with the article in the bioethics journal, ethicists point out that the best success rate for large organ transplants is about 84%. This means that one in five face transplants could fail, leading to the loss of the transplanted face. Failed transplants would leave vulnerable people much worse off, some argue, because the operation would have removed whatever facial tissue the patient had before the procedure.

Critics also contend that hand transplants, which the Louisville team has performed,



Helping hand: experiments with other complex organs point the way to full facial transplants.

have not been successful enough to justify further similar experiments. And they point out that immunosuppressive drugs given to face-transplant recipients have toxic side-effects, which could endanger the recipient's life.

"I think there hasn't been enough background research to prove that it will work, and they haven't figured out what to do if it doesn't work," says medical ethicist Arthur Caplan, director of the Center for Bioethics at the University of Pennsylvania in Philadelphia.

The transplant team counters that the hand transplants have shown how successful large organ transplants can be, because only two of 24 hands transplanted around the world have been rejected. However, this sample is small and the grafts have all been performed relatively recently. The team has also conducted numerous preparatory experiments, such as face transplants from one cadaver to another.

Barker says that careful patient selection would address the concerns about a failed transplant. The team would try to select a recipient recovering from a recent disfiguring event instead of one who has already gone through years of reconstructive surgery. That way, if the graft failed, the patient would have lost little. Careful preparation could also help the recipient deal with the inevitable media storm that would follow such a transplant, the Louisville team says.

Critics counter that the risk of the procedure is just too high. Any day now, the Netherlands review panel is expected to decide who it thinks is right.

US health agency opens landmark clinical centre

Helen Pearson

Clinical research at the National Institutes of Health (NIH) was due to receive a major boost this week with the official opening of a 242-bed, \$540-million centre at the agency's main campus in Bethesda, Maryland.

A ribbon-cutting ceremony on 22 September launched what is said to be the largest centre of its kind in the world. It is named the Mark O. Hatfield Clinical Research Center, after an Oregon senator who strongly supported the NIH.

The centre will form the mainstay of the NIH's intramural clinical research programme at a time when the institute's director, Elias Zerhouni, is striving to improve links between laboratory and clinical research. "There will be a lot of people watching — and hopefully we won't disappoint," says the centre's director, immunologist John Gallin.

Laboratories and patient wards sit side by side in the new building and most are equipped for either function, allowing space to be transferred seamlessly from one purpose to the other.

The centre also boasts special facilities, such as a 12-bed unit devoted to clinical studies of obesity. To take account of its residents' unusual size, the unit includes an enlarged magnetic resonance imaging scanner, widened doorways, reinforced toilets and vending machines that log the calorific value of patients' selections. One of the issues researchers plan to study is whether obese people's physiology and brain activity change after stomach-reducing surgery.

But, like other clinical research programmes in the United States, the centre faces a challenge in attracting staff with appropriate experience, as most physicians are encouraged to pursue lucrative work on patients rather than research. The centre is developing a variety of measures to entice researchers, including a computer system to cut the paperwork surrounding clinical trials.

The centre will also focus effort on areas that are sometimes ignored by other medical institutions, such as research into rare diseases, says the president of Boston's Dana-Farber Cancer Institute, Ed Benz, who chaired a panel earlier this year on the future of NIH intramural clinical research.

For more news and analysis go to
news@nature.com
www.nature.com/news

Students set up forum to debate hot topics

Jonathan Knight, Seattle

Graduate students at the University of Washington in Seattle are leading an initiative to encourage dialogue between scientists and the public on hot research issues.

Five biology students at the university set up the Forum on Science Ethics and Policy (FOSEP), which aims to stir up public discussion on areas of research that raise ethical concerns, such as human embryonic stem-cell research.

The forum will hold its first major public meeting — in a hall that seats an audience of 700 — on 18 October, when a panel of scientists will discuss stem-cell research. The meeting falls just before the presidential election, in which stem cells have emerged as an important issue.

Melanie Roberts, a neurobiology graduate student and one of FOSEP's co-founders, says the project sprang from concern that there weren't enough opportunities for scientists and the general public to exchange views on the controversial issues surrounding research, from genetically engineered crops to nanotechnology.

In January, she teamed up with four other policy-minded classmates to start FOSEP. They contacted numerous experts for advice, raised funds, drew up a budget and scheduled events.

FOSEP has an unconventional goal. Whereas other public outreach efforts tend to emphasize science education for a lay audience, FOSEP is intended as much for the benefit of scientists, who may have a limited grasp of how the public sees their work.

"We don't want to talk about educating the public," says Roberts. "We are not promoting a particular position. We want to provide a forum for the public and the scientists to communicate."

Matt Zonarich, national coordinator for



Fount of knowledge: students at the University of Washington want to inspire public discussion.

the Joint Steering Committee for Public Policy, a biomedical research advocacy group based in Washington DC, notes how unusual this is. "I've been doing grassroots biomedical research advocacy for almost four years, and I have yet to come across another effort with the same mission," he says.

The October event, which is being promoted through local newspapers and radio shows, will include a panel of two stem-cell researchers, a bioethicist and an expert on health policy. For most of the evening the panel will respond to questions from the audience, moderated by Seattle public radio host Ross Reynolds. A web-cast will be available and the speakers will continue to answer questions online after the event.

Research policy experts welcome the forum approach. "Roberts and others are thinking that maybe not all scientific research is good, or should be done," says

Jane Maienschein, director of the Center for Biology and Society at Arizona State University in Tempe. "They are learning to be more sympathetic to competing views." Maienschein gave a FOSEP-sponsored public lecture in May.

The intention is to hold at least one forum a year, Roberts says. As the founding members of FOSEP graduate — Roberts intends to finish in May — they will be replaced by younger students. To make the organization sustainable in the face of student turnover, FOSEP is applying for grants and drawing up a permanent organizational structure.

Roberts says she was initially concerned that no one would take a student group seriously. But as it turns out, FOSEP's student leadership has been an asset. "I think a lot of people have opened their doors to us as graduate students who might not have for faculty," she says. ■

Research plane will scale uncharted heights

Quirin Schiermeier, Munich

The German research ministry has approved the construction of a high-performance research aircraft that should be ferrying European atmospheric researchers into the upper atmosphere by 2007.

The €67-million (US\$81.5-million) High Altitude and Long Range Research Aircraft (HALO) will be able to fly 8,000 kilometres before refuelling, allowing scientists to monitor conditions in some of the crucial regions for climate research around the world — from Brazil to the South Atlantic, and from southeast Asia to arctic Russia. Its maximum payload of more than 3 tonnes will

enable several groups to conduct experiments simultaneously during one flight.

The plane is designed to reach unprecedented heights of 16,000 metres — the upper part of the boundary region between the stratosphere and the troposphere. Other research planes can reach only a maximum of 13,000 metres. Studying this higher region will improve researchers' understanding of atmospheric circulation. It will also help them to evaluate whether pollution from the growing fleet of commercial aircraft is having an impact on cloud formation and climate.

HALO will be based on a Gulfstream

G550 — a luxury passenger jet. But its interior will bear little resemblance to its plush origins. The seats will be replaced by two rows of cabin-high instrument racks, with small niches for up to six operators. A window in the plane's floor will allow observation of the atmosphere below.

The United States also has a high-specification research plane in the works. Its High-performance Instrumented Airborne Platform for Environmental Research (HIAPER), commissioned three years ago, is due to fly its first missions next year. But the German plane is based on a newer, larger jet. ■

Study warns of 'avoidable' risks of CT scans

Jim Giles, London

Hundreds of fatal cases of cancer are being induced in children every year by hospital scans that could be replaced by other diagnostic methods, researchers have claimed.

The researchers' concern centres on the use of computed tomography (CT) scans; these create images of internal organs by combining X-rays taken from different angles. An initial analysis by David Brenner, a radiation biologist at Columbia University in New York, suggests that a single year of paediatric CT scans will ultimately cause 2,500 cancer deaths in the United States alone. But a survey of radiologists suggested that a third of these scans could be avoided, adds Brenner, who presented his work at an international conference in London — "Children with Leukaemia: Incidence, Causal Mechanisms and Prevention" — on 7 September.

The use of childhood CT scans has mushroomed since 1990, when a new generation of devices cut the scanning time from 1 minute to 1 second. The shorter scans have been a boon to clinicians, who use them to study

dangerous conditions such as appendicitis in younger and less cooperative children.

Yet the technique exposes patients to tens of times more radiation than conventional X-ray examinations. Brenner says the doses are comparable to those received by people around two kilometres away from the atomic bomb blasts at Hiroshima and Nagasaki in 1945. Extrapolating from the bomb studies, he says that a single CT scan on a newborn girl has a 1 in 800 chance of causing fatal cancer.

Debate about the safety of CT has intensified in the United States this year. Brenner, for example, has highlighted the risks of adult full-body scans (D. Brenner and C. D. Elliston *Radiology* **232**, 735–738; 2004). But he and Elaine Ron, an epidemiologist at the National Cancer Institute in Bethesda, Maryland, who worked with him on the paediatric studies, say the risks to children are only just beginning to be quantified.

Brenner says that the benefits of appropriately using the technique almost always outweigh the risks, but adds that many scans could be avoided. A straw poll taken at a

2001 radiology conference, held in Chicago, Illinois, showed that delegates believed that up to 30% of paediatric scans were unnecessary (*Pediatr. Radiol.* **32**, 242–244; 2002).

The US Food and Drug Administration, which regulates medical devices, already advises clinicians to use lower radiation doses in children. But a 2001 survey of practice in one US hospital indicated that radiologists may not be doing this (A. Paterson, D. P. Frush and L. F. Donnelly *Am. J. Roentgenol.* **176**, 297–301; 2001). Kieran McHugh, a paediatric radiologist at Great Ormond Street Hospital for Children in London, says UK clinicians are becoming more aware of the issue. "But I still see images taken at adult settings," he adds. "You could use a quarter of the dose."

A proposed epidemiological study by researchers at the University of Newcastle upon Tyne, UK, may provide better data on the risks. Mark Pearce and Louise Parker hope to study the hospital records of 100,000 children who received scans. But Pearce cautions that several such studies will be needed to properly assess the risk of CT scans. ■

Digital mastery peers through cracks at early Picasso

Barbara Simm

For almost a century, Pablo Picasso's early painting *Rue de Montmartre* held a surprising secret. But a few years ago, X-ray analysis showed that beneath its restrained portrayal of some hunched-up figures there lurked a very different painting of the decadent side of Parisian nightlife.

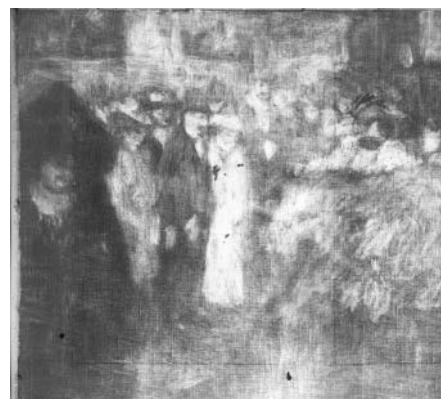
Now William Shank, former chief conservator at the San Francisco Museum of Modern Art, has digitally recreated the hidden work in all its colourful glory. Named simply *A Hidden Picasso*, it is being exhibited for the first time this month at the Guggenheim Museum Bilbao in Spain, to coincide with an international meeting of art conservationists taking place in the city.

Shank began with the shadowy, black-and-white image revealed by the X-rays — the method used routinely by conservators to see through thin layers of paint. He then identified the painting's original colours, using cameras to peer through the networks of tiny cracks in *Rue de Montmartre* and computer imaging techniques to build up an image of what lies beneath it.

Experts date the uncovered painting to 1900, which would make it the first work that the artist painted after arriving in Paris from Barcelona at the age of 19 — giving the elusive work special significance for art historians. ■



Buried treasure: beneath Picasso's *Rue de Montmartre* (left) lies the first picture the artist painted after arriving in Paris. X-rays revealed *A Hidden Picasso* (below, left), which has now been digitally enhanced to display its original colours.



: ON ME

IMAGES: SFMO/MA

Heads roll at nuclear lab following inquiry into lost disks

Washington More than ten employees at Los Alamos National Laboratory in New Mexico have been fired, demoted or reprimanded following security scandals.

Laboratory officials have spent the past two months investigating the disappearance of computer disks containing classified information from a wall safe at the lab. Such security problems sparked a slew of local and federal security inquiries that have left researchers at the lab idle since mid-July (see *Nature* **430**, 387; 2004).

Four employees, whose names were not released, were dismissed in connection with the disappearance of the disks and an unrelated lab accident in which an intern was partially blinded by a laser, according to University of California spokesman Chris Harrington. A fifth employee resigned.

Of the 18 other employees put on investigatory leave, seven were reprimanded, demoted or received pay cuts, and ten returned to work after being exonerated. One remains on leave pending further investigations.

The lab is expected to resume normal operations by mid-October.



A small group of dugongs near Okinawa may be endangered by a planned US military base.

Conservationists blast plans for offshore army base

Tokyo Japan's defence agency has begun to drill into the seabed off the island of Okinawa in preparation for an offshore military base for US forces. But protesters have bombarded both the US president and Japan's prime minister with calls to abandon the base, saying the construction will have a devastating effect on local sea life.

The site, off the city of Henoko, plays host to coral reefs that serve as feeding grounds for a dwindling population of dugongs (*Dugong dugon*) — large mammals, also known as sea cows, that are listed as vulnerable to extinction. A representative of the Okinawa-based Save the Dugong Foundation says that the population near

the island might contain fewer than ten animals. "The dugong is symbolic of the various kinds of marine life that would be lost," she says. A group of more than 400 conservation organizations is calling for the construction project to be abandoned.

The defence agency says that the current drilling will contribute to a study of the environmental impact of the proposed base. Officials say the station is needed to decrease the burden of US forces on Okinawa itself, where residents have long protested the military presence.

Scripps stays wedded to original Florida site

San Diego The Scripps Research Institute in La Jolla, California, last week decided to go ahead with its plans to build a research facility at a controversial site in Florida.

The decision, made on 13 September, calls for the establishment of a centre for biomedical research and drug design on a 900-hectare site near a wildlife refuge in Palm Beach County.

Environmentalists oppose the location, which they say is on ecologically sensitive wetlands. "We will continue to pursue all legal options to challenge the development," says Manley Fuller, president of the Florida Wildlife Federation.

County and state officials plan to spend about \$750 million in coming years to develop a science community in Florida, with Scripps as its centrepiece.

US advisers get tough on antidepressants

Washington Advisers to the US Food and Drug Administration (FDA) have recommended that it should label all antidepressants with a sweeping warning that the drugs cause some children to attempt suicide.

After a hearing last week, the committee voted 15 to 8 to advise the FDA to issue a 'black box' warning — the strongest possible — about the drugs. This is a big step-up from the warning labels adopted by the FDA in March, which say that antidepressant drugs are associated with a risk of suicide in children, but do not necessarily cause it.

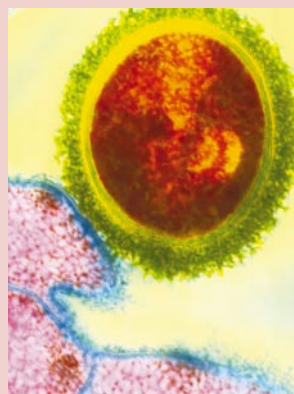
The FDA usually adopts the committee's recommendations, but will discuss the issue internally before announcing a decision, says Robert Temple, director of the agency's Office of Drug Safety.

If the advice is followed, the new warning will appear on all antidepressants on the US market. Drug companies will also have to post warnings prominently in any television or magazine advertisements.

Flesh-eating bacteria get a taste of success

Munich The harder you look, the more you find ... and that isn't always good news. A European research network has found that infections with group A *Streptococcus* (see right), also known as 'flesh-eating bacteria', are more common than previously thought.

Scientists in 11 countries, led by microbiologist Aftab Jasir at Lund University in Sweden, found five times as many such infections as they anticipated in the first 18 months of their surveillance. They estimate that about 20,000



severe cases occur in the European Union each year, and this seems to be increasing. "In Britain alone, the number of reported cases has doubled in the past five years," says Jasir.

Most infections with group A *Streptococcus* are benign. But the bacteria can chew through tissue, muscle and blood vessels, causing death in 20–30% of cases.

Amid fears that the bacteria are also developing resistance to antibiotics, scientists are pushing for more research into their basic biology.

President of Whitehead Institute steps down

Washington The first woman to head the prestigious Whitehead Institute for Biomedical Research has announced that she will resign on 1 November to return to full-time research.

Susan Lindquist became director of the 22-year-old centre in Cambridge, Massachusetts, in 2001. She presided over a major downsizing of the institute as its

genomics work, led by Eric Lander, was transferred to the new Broad Institute, also in Cambridge.

During her tenure as president, Lindquist's research into prions flourished. "Her work has always been highly original and impressive and, if anything, that has increased in pace," says Maxine Singer, chair of the Whitehead board. Singer says the institute will begin a search for a permanent director, and will name an interim director before 1 November.

Deckchair science

There are worse places to do research, but for the Caribbean cruise ship *Explorer of the Seas*, home to two climate laboratories, it's not all smooth sailing. Emma Marris finds out why.

Oceanographer Rik Wanninkhof was stranded. Working at the Atlantic Oceanographic and Meteorological Laboratory in Miami, Florida, he had developed an instrument to measure carbon dioxide levels in the ocean. After a number of dry runs, he was eager to test it at sea, and so he placed several of his sensors on the only research vessel available. But the ship was scheduled to sail to South America and Europe, so if anything went wrong, he would have to fly a long way at considerable expense to fix the problem.

Salvation came from an unexpected quarter: Royal Caribbean Cruises. The holiday company offered him the chance to install his sensors on *Explorer of the Seas*, a 3,000-passenger luxury cruise ship, which is also home to two sophisticated laboratories, one each for atmosphere and ocean research. So Wanninkhof's instruments went on an extended tour of the Caribbean.

The cruising labs are a joint project between Royal Caribbean and the Rosenstiel School of Marine and Atmospheric Science at the University of Miami. As research vessels are both expensive and scarce, oceanographers have been jumping at the chance to place instruments on board *Explorer of the Seas* for free. They may also be attracted by the stability of a large ship, as opposed to the pitch and heave of a tiny, crowded research vessel, where even the saltiest of scientists go green about the gills. Whatever the reason, they have been signing up in their droves.

But it's not all plain sailing. A cruise ship that only stops in port and never leaves the Caribbean is not suitable for every project. And a notable lack of publications based on data from the ship has some researchers questioning the usefulness of the arrangement. This dearth of results combined with the declining fortunes of the cruising industry



Lap of luxury: the cruise liner *Explorer of the Seas* doubles as a floating lab for climate research.

has left the project with an uncertain future.

The idea of cruise ship labs originated as part of Royal Caribbean's effort to clean up its image following a series of pollution scandals in the 1990s. After admitting that its ships had been illegally discharging oil and waste chemicals into the sea for years, the company promised to stop and agreed to pay millions of dollars in fines. In 1996, it established the Ocean Fund, which makes grants to conservation organizations.

To run it, the firm turned to Otis Brown, dean of the Rosenstiel School. One day, Brown happened to complain in front of Royal Caribbean's president, Jack Williams, about the cost of research vessels. What emerged was a realization that cruise ship labs could benefit both the cruise line and oceanography.

Sea view

For researchers, the rewards were obvious. Cruise ships sail the same course over and over, so long-term data sets can be assembled about certain locations. And they are massive — more like reclining skyscrapers than boats — so their inherent stability means that the instruments don't sway or get knocked about. Perhaps most importantly, researchers get more of a scarce resource: sea time.

Royal Caribbean was expected to benefit too. Brown and Williams imagined that hosting a lab investigating such things as climate change and the transportation of invasive species in bilge water would help the company's environmental image. The lab was also expected to be an attraction for passengers, on a par with the ice-skating rink and climbing wall.

Additionally, as part of the deal, Brown agreed to find one researcher each week to go on board for the benefit of the cruisers. Each

visiting scientist would give two lectures, a general talk called 'Dive into marine science' and a non-technical explanation of their own area of expertise. They would also give tours of the labs and act as an informal ambassador of science in their free time, if only by chatting with passengers over drinks.

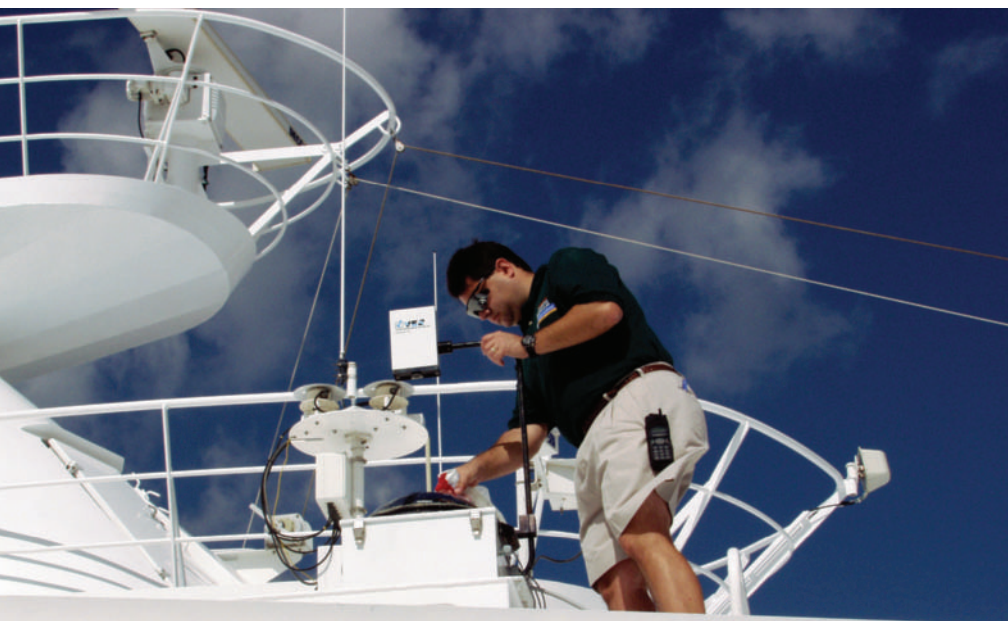
For its part, Royal Caribbean came up with US\$3 million to add the two laboratories to *Explorer of the Seas* as it was being built in Finland for launch in 2000. To fit the labs with the latest equipment, Brown convinced several initially sceptical federal research agencies, including the National Oceanographic and Atmospheric Administration (NOAA), to chip in a few million more.

It wasn't hard to find researchers willing to take advantage. Strapping instruments to boats other than dedicated research vessels has long been a tactic of climate scientists, one that has allowed them to get information from around the world, often from ships that ply repeated tracks. But the instruments on such ships are often unmanned and cheap, even disposable.

Star attraction

In contrast, the labs on *Explorer of the Seas* are attended continually by two onboard technicians. The instruments, which are either bolted to the hull or tucked away on a mast, are top of the range. They include two acoustic Doppler current profilers to image the currents and marine life at various depths, an emitted radiance interferometer to measure the temperature of the sea's surface with extreme accuracy, and a radian wind profiler, which tracks wind speed at various elevations overhead. One machine does nothing but take pictures of clouds.

The labs themselves, conceived by Royal



Cruise control: from the labs on the liner, researchers guide a range of instruments and experiments.

Caribbean designers, look as if they have come straight off a Hollywood set. They combine flat-screen monitors, hidden wires, blue neon lights and burnished chrome in an aesthetic statement Brown describes as somewhere between Jules Verne and *Star Trek*.

The boat has two routes, alternating weekly. One track goes from Miami to San Juan, Puerto Rico, and the other makes a loop around Cuba, stopping at Jamaica and Cozumel off the coast of Mexico. As the ship ploughs through the same waters, the instruments collect and transmit the data back to the Rosenstiel School via satellite.

This sort of repetitive sampling is particularly well suited to climate research, the main thrust of most of the work on *Explorer of the Seas*. The researchers — from the University of Miami, NOAA and an increasing number of other institutions — have set out to collect data on as many atmospheric and oceanic variables as possible to learn how the air, water, wind, sun and sea life interact, as well as how the whole system changes throughout

the year. Ultimately, the data generated should help them to understand how long-term climate trends such as El Niño and global warming are affecting the region.

Choppy waters

But it's a slow business, and very few papers have come out of the research so far. This has led some researchers to wonder whether the flashy labs will produce interesting science. "They have been slow to work it into a scientifically usable shape," says Tom Rossby, an oceanographer at the University of Rhode Island in Kingston, who puts his equipment on freighters, although he concedes that such long-term studies could take more time to come to fruition.

One reason for the delay may be that the ship is always in motion, making it impossible to lower deep-water sampling devices, says John Orcutt, deputy director for research at the Scripps Institution of Oceanography in La Jolla, California. As a result, the water samples all come from the top layer of the sea, a

comparatively hurly-burly place where long-term trends are harder to pick out. "To get at climate change, you sample deep," he says. "Their signals for things such as long-term change are going to be very noisy." He expects that, because of this, real results will take longer to emerge from the background din.

Lack of publication is rarely considered a positive sign by grant makers, and Brown spends a good deal of time cobbling together funding from a variety of agencies and private donors. After grants from NOAA and the National Science Foundation expired, for instance, Brown succeeded in winning new support from the Office of Naval Research and the University of Miami.

Turning tide

Compounding the money problem, the cruise line itself has withdrawn some of its support. Although it still provides the free berth and lab space for researchers, it no longer helps to maintain the instruments or cover the salaries of the onboard technicians. Brown says that the change was abrupt, following the drop in tourism after the terrorist attacks on 11 September 2001: "They said: 'We're going to try to keep the ship running. Your problems are your problems'."

Of course it's not new for long-term studies of the type done on the ship to face funding challenges, Brown points out. "The ocean-science support system has not really been sensitive to the needs of long-term projects," he complains. Dean Roemmich, a professor at Scripps, says that funding such studies is a constant challenge. "It takes a lot of patience to see results," he says. "It's hard for people to understand. There is this period where we are trying very hard to prove our value."

On the bright side, the results that might prove the project's worth are finally emerging. In February, Wanninkhof's tests led to one of the first published findings from *Explorer of the Seas*. His instruments picked up a significant seasonal variation in carbon dioxide levels that correlated with temperature changes, despite expectations that CO₂ levels were fairly constant in the region. This finding might make it possible to estimate CO₂ levels by taking the temperature of the sea surface from a satellite (A. Olsen, J. A. Triñanes and R. Wanninkhof *Remote Sens. Environ.* **89**, 309–325; 2004).

Despite all the uncertainties, researchers are queuing up to board *Explorer of the Seas* — the waiting list is eight months long. It's quite a change for those used to rougher outings. When not lecturing or installing instruments, they can swim, sunbathe or go ice-skating. But it's no junket. They can only drink outside working hours, and are not allowed to gamble. "This is not a 'lie by the pool and get a drink whenever you are thirsty' kind of thing," Brown says. "You actually do work."

Emma Marris is an intern in Nature's Washington office.

♦ www.rsmas.miami.edu/rccl

By chance, or by design?

Enzymes are well known for speeding up reactions. But have they evolved to use quantum mechanics to exert their effects? Philip Ball meets the researchers who are trying to find out.

Most biologists would scoff at the idea that their subject is simply applied quantum mechanics. But for some enzymes — the catalysts of biology — quantum effects may be an important part of the way they work. This revelation has left chemists and biologists arguing about whether enzymes have evolved to do this, or whether the effect would happen regardless of the enzymes' activity.

At the heart of this debate is the issue of how much control enzymes have over the processes that affect their catalytic power. It's a debate about the limits of molecular biology's capabilities and inventiveness. For researchers trying to design new enzymes for chemical and biochemical synthesis, the idea that natural enzymes can manipulate quantum effects to their advantage could suggest new possibilities. At the same time, it could make the design process a lot more complex.

The effect in question is known as quantum tunnelling. This offers quantum objects — such as a hydrogen atom — a way to cross an insurmountable energy barrier that they cannot get over 'classically'. Rather than needing sufficient energy to overcome the barrier, the object just tunnels through it instead. "There is now ample evidence that tunnelling occurs in enzymes, and the biochemical community generally accepts this to be the case," says Nigel Scrutton, a biochemist at the University of Leicester, UK. In fact, tunnelling can speed up a reaction by a thousand times or more, he adds.

But the question of whether enzymes have evolved to make the most of quantum tunnelling has provoked a heated reaction. Not everyone agrees that molecular biology is smart enough to have such an effect.

"This is a big red herring," says Arieh Warshel, a theoretical biochemist at the University of Southern California. The critics say that quantum tunnelling happens regardless of what the enzyme does, and has nothing to do with evolutionary fine-tuning.

In classical terms, quantum tunnelling is like a marble inside a jar vanishing and then reappearing outside the jar. It seems impossible but, in the quantum world, it can happen because all quantum particles have wave-like properties. This means that their location is defined by a varying probability over space —

so there is always a small chance that the particle will appear on the other side of a barrier.

The probability that quantum tunnelling will occur falls as the object gets heavier, which is why it never actually happens to a marble. In fact, it is unlikely to occur for most atoms in chemical reactions — unless they are as small and light as a hydrogen atom. "Hydrogen transfer is probably one of the most fundamental and prevalent processes in biology," says Judith Klinman, a protein chemist at the University of California, Berkeley, who has championed the case for tunnelling in enzymes since the late 1980s.

Digging in

The possibility that hydrogen tunnelling occurs in biochemical reactions was first raised over 30 years ago. But it wasn't until 1989 that Klinman and her colleagues reported direct evidence for it¹. They studied a yeast enzyme called alcohol dehydrogenase (ADH), which transfers a hydrogen atom from an alcohol to a small molecule known as nicotinamide adenine dinucleotide.

To test for tunnelling, the team looked for the kinetic isotope effect (KIE) — a small change in the rate of the enzyme's reaction. If the hydrogen atom to be moved from the alcohol is switched for a heavier isotope — deuterium or tritium — the bond energy changes,

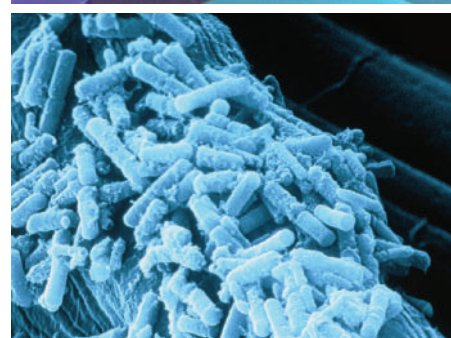
which in turn alters the reaction rate. Classically, it is fairly easy to predict what this change should be. But if tunnelling is involved, then there is an additional effect on the rate. So a KIE with properties different from the one predicted classically indicates that quantum tunnelling has taken place. This is exactly what Klinman and her team saw.

Since then, they and others have found evidence for tunnelling in several enzymatic reactions^{2,3}. Klinman now thinks that enzymes can enhance hydrogen tunnelling to increase their reaction rates. Working with chemists in Italy in 1999, she found that some enzymes seem to adjust the amount of tunnelling to suit their operating temperature⁴.

This time, the team was looking at an ADH from the thermophilic bacterium *Bacillus stearothermophilus*. This enzyme works at a temperature of 65 °C, some 40 °C hotter than the yeast equivalent. Tunnelling is expected to make less of a contribution to the

"Tunnelling is a fact of life, but life has no special effect on tunnelling."

— Willem Siebrand



Bacillus stearothermophilus uses an enzyme that seems to adapt to its working temperature.

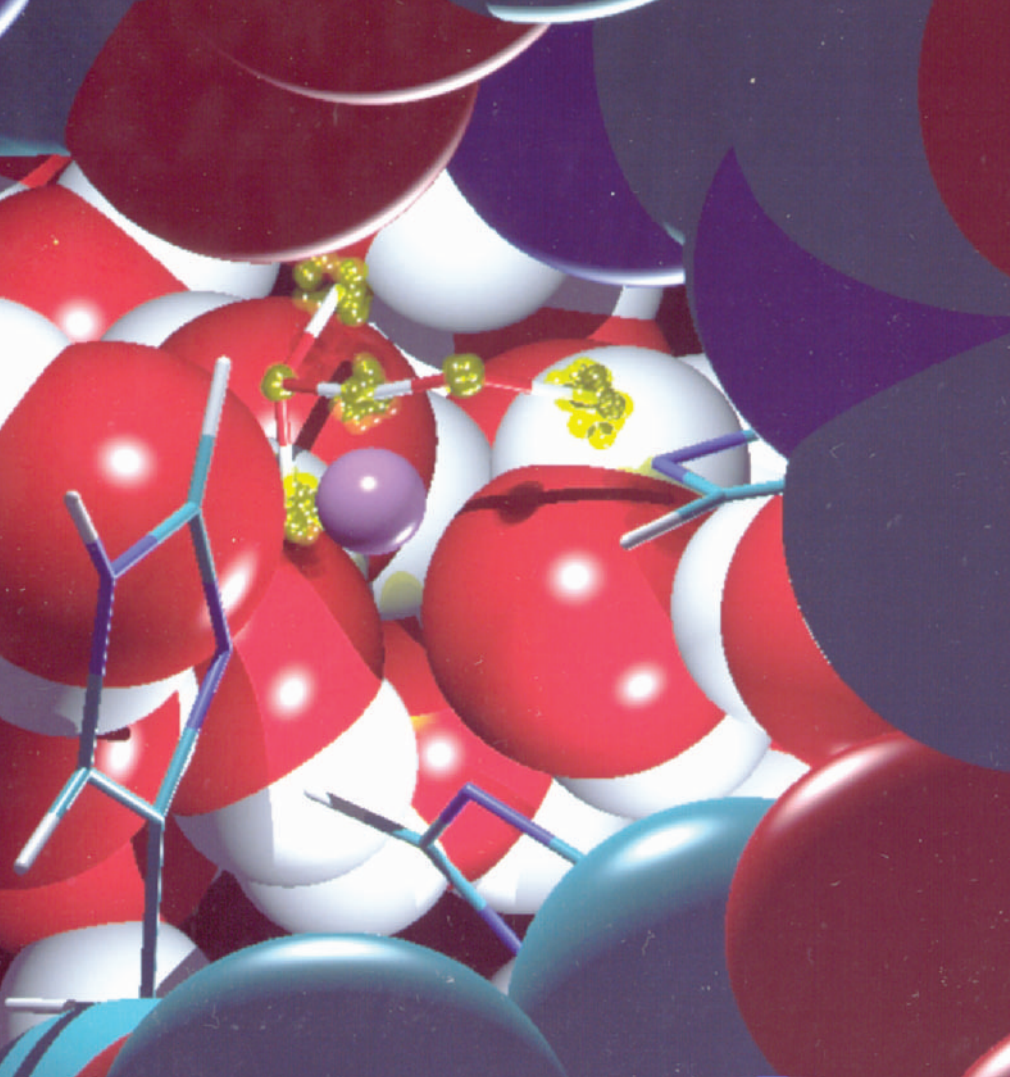
reaction rate for enzymes operating at higher temperatures. But the researchers found that the tunnelling contribution was similar for both the yeast and the thermophilic ADH.

Tunnel vision

Others argue that tunnelling occurs irrespective of what the enzyme is doing. "All hydrogen reactions involve some degree of tunnelling, depending on the temperature," says Richard Finke, a chemist at Colorado State University in Fort Collins. So perhaps biology isn't 'using' this quantum effect, but simply failing to do anything either to prevent or to enhance it.

Finke says that a proper test of the hypothesis would mean comparing the amount of tunnelling that occurs in the pres-

M. GESPL



Models of biocatalysis suggest that quantum tunnelling occurs whether or not an enzyme is present.

ence and absence of the enzyme. But this is difficult to do, because in most cases the reaction won't proceed without an enzyme.

In 2000, Finke found the test he was looking for. The enzyme methylmalonyl-CoA mutase catalyses a hydrogen-transfer reaction but requires a cofactor, a molecule called AdoCbl, for the reaction to proceed. The enzyme breaks a cobalt-carbon bond in AdoCbl to generate a free radical, which then pulls a hydrogen atom off another molecule (methyl malonate-CoA). Previously, this reaction had shown evidence of tunnelling.

But a very similar reaction occurs without the enzyme if AdoCbl is heated in ethylene glycol solution. In the enzyme-free reaction, the ethylene glycol, rather than methyl malonate-CoA, provides the hydrogen atom. But Finke and his colleagues figured that the reaction rate should be largely unaffected by these different sources of hydrogen, so that the extent of tunnelling with or without the enzyme could be directly compared.

They found that the tunnelling rates were more or less identical, within experimental error^{5,6}. In other words, there is no reason to believe that the enzyme is doing anything special. Such results have persuaded Willem

Siebrand, a theoretical chemist at the Steacie Institute for Molecular Sciences in Ottawa, Canada, that "tunnelling is a fact of life, but life has no special effect on tunnelling".

But, as Finke points out, the evolutionary pressure on the enzyme to take advantage of quantum-mechanical tunnelling is weak in this case. This is because the enzyme's main job is to break the cobalt-carbon bond in AdoCbl, rather than to pluck out a hydrogen atom. The different hydrogen sources for the two reactions also worry Klinman and Scrutton. "It is very dangerous to generalize on the basis of one comparison of model chemistry with enzymatic chemistry," says Scrutton.

Warshel argues that because it is so hard to compare catalysed and uncatalysed reactions experimentally, it is better to use computer simulations. In 1996, he and Jenn-Kang Hwang from National Tsing Hua University in Taiwan simulated the behaviour of the enzyme carbonic anhydrase and incorporated quantum effects such as tunnelling⁷. They compared their model with one for the same process happening in an enzyme-free solution, and found a similar amount of tunnelling in both cases.

It is much harder than some people think for enzymes to manipulate their reactants

to optimize the amount of tunnelling, says Warshel. To do this, they would have to change the shape of the energy barrier that separates the reactants from the products. Warshel thinks that enzymes are too floppy for that. "The idea that enzymes can do 'anything' is wrong," he says.

But those who believe that enzymes can make use of quantum tunnelling argue that the proteins could use dynamic behaviour — such as their molecular vibrations — to 'squeeze' the hydrogen through the energy barrier. Scrutton admits that this idea remains controversial, however.

Good vibrations

Klinman claims that an enzyme's vibrations can be used to make tunnelling easier — for example, by reducing the distance over which the hydrogen has to tunnel between the source and the target molecules, or by altering the relative energy levels of the reactants and products. In theoretical calculations⁸, the hydrogen-tunnelling distance in a reaction catalysed by the enzyme soybean lipoxygenase seems to be shorter than the distance between the source and target molecules. This suggests that vibrations are needed to bring the two molecules closer together when tunnelling occurs.

Klinman is convinced that evolution can select particular vibrational states to enhance enzyme catalysis. Scrutton, meanwhile, believes that even if vibration-assisted tunnelling does occur, it is not absolutely essential. In his view, the available evidence suggests that many enzymes are already optimized for tunnelling without the help of vibrations. For example, his group has found that the enzyme trimethylamine dehydrogenase doesn't seem to exploit vibrations in its natural form but does make use of them in a mutant form⁹.

Others dispute the whole idea. Siebrand thinks that "proteins are far too flexible to be good squeezers", and that the process would require too much energy. Warshel agrees. "On balance there is no evidence for dynamical contributions to catalysis," he says.

The debate shows little sign of being resolved quickly. And until it is, we must remain uncertain about the limits of nature's ingenuity. ■

Philip Ball is a consultant editor for Nature.

1. Cha, Y., Murray, C. J. & Klinman, J. P. *Science* **243**, 1325–1330 (1989).
2. Kohen, A. & Klinman, J. P. *Acc. Chem. Res.* **31**, 397–404 (1998).
3. Sutcliffe, M. J. & Scrutton, N. S. *Eur. J. Biochem.* **269**, 3096–3102 (2002).
4. Kohen, A., Cannio, R., Bartolucci, S. & Klinman, J. P. *Nature* **399**, 496–499 (1999).
5. Doll, K. M., Bender, B. R. & Finke, R. G. *J. Am. Chem. Soc.* **125**, 10877–10884 (2003).
6. Doll, K. M. & Finke, R. G. *Inorg. Chem.* **42**, 4849–4856 (2003).
7. Hwang, J.-K. & Warshel, A. *J. Am. Chem. Soc.* **118**, 11745–11751 (1996).
8. Hatcher, E., Soudackov, A. V. & Hammes-Schiffer, S. *J. Am. Chem. Soc.* **126**, 5763–5775 (2004).
9. Basran, J., Sutcliffe, M. J. & Scrutton, N. S. *J. Biol. Chem.* **276**, 24581–24587 (2001).

School arrays benefit science as well as students

While cosmic rays arouse interest in physics, more rare phenomena may be spotted.

Sir — I read with interest your Editorial on “Schools at 10^{20} eV and beyond” (*Nature* **429**, 685; 2004). We applaud the University of Nijmegen physicists for their success with the high-school project on astrophysics research with cosmic rays (HISPARC), which has installed cosmic-ray detectors at high schools across the Netherlands.

As the detection rate for these rare high-energy cosmic rays is in the order of one a year per square kilometre of collecting area, it is important to increase the collecting area (by increasing the number of schools involved), and so we welcome HISPARC to a growing family of similar projects across North America and Europe.

In addition, these arrays can search for coincident, nonrandom ‘hits’ across a large area that could result from correlated primary events from some of the highest-energy objects in the Universe. Again, such phenomena are expected to be very

rare and would benefit from the largest collection area possible.

The first cosmic-ray project with an educational dimension was Canada’s ALTA project, ‘a search in Alberta over a Large area for cosmic-ray shower Time coincidences using an Array of detectors’. This started in 1996 and the idea has been spreading across the world ever since. The ALTA network now consists of 15 sites in high schools across Alberta covering an area of some 100,000 square kilometres.

In addition to searches for cosmic-ray phenomena, the ALTA collaboration of high-school students, teachers and university researchers are pursuing investigations into atmospheric effects on the detection rate for high-energy cosmic rays. Additionally, we will be investigating the use of the arrays to monitor the soft cosmic-ray flux from the Sun — to provide an online ‘solar weather report’.

The North American projects are grouped together as the NALTA collaboration (<http://csr.phys.ualberta.ca/nalta>) and now cover a total collecting area of a few hundred thousand square kilometres. Similar projects in Europe include the Stockholm Educational Air Shower Array (SEASA) in Sweden and Sky-View in Germany. In addition, the Italian Extreme Energy Events (EEE) project is just getting started. A project in the Czech Republic is being jointly developed by the Czech Technical University in Prague and the University of Alberta.

Clearly, more arrays mean more opportunities for students and for science. We would like to welcome HISPARC to this family and wish it continuing success for the future.

James L. Pinfold

Department of Physics, University of Alberta,
Edmonton, Alberta T6G 2N5, Canada

Hopes remain for an Alzheimer’s vaccine

Sir — Your News story “Doctors seek lost data on Alzheimer’s vaccine” (*Nature* **430**, 715; 2004) correctly points out that Elan’s approach to treating Alzheimer’s disease is “one of the few potentially viable options for controlling the disease”. However, we would like to correct an inaccurate statement regarding the results of the discontinued AN-1792 trial.

Research carried out by Nick Fox of the Institute of Neurology in London, presented at the International Conference on Alzheimer’s Disease and Related Disorders in July, measured the brain volume of patients enrolled in Elan’s clinical trial of AN-1792. Fox found that the brain volume of patients treated with AN-1792 who responded to the treatment by making antibodies diminished by 3% — not the 6% stated in your News story — and this decrease was only 1% greater than the brain-volume reduction in those who were treated with a placebo.

In addition, Fox commented that, although the brain-volume reduction was unexpected (not “alarming”, as the News story suggests), the small changes seen might be due to the removal of plaque lesions in the brain tissue of patients suffering from Alzheimer’s.

We continue to pursue a number of immunotherapy approaches to Alzheimer’s disease, including several in preclinical phases of development, with the goal of bringing a meaningful treatment to the

millions of individuals, families and caregivers affected by this devastating disease.

Dale Schenk

Elan Corporation, 800 Gateway Boulevard,
South San Francisco, California 94080, USA

German reforms would make little difference

Sir — Your Editorial “Weak at the centre” and News story “Court ruling upsets hopes for career reforms” (*Nature* **430**, 593 and 599; 2004) describe the recent failure to replace the traditional career path at Germany’s universities, the *Habilitation*, with a US-style junior-professor system.

It is true that most young German scientists are unhappy with the Supreme Court’s decision to stop the reform by declaring it unconstitutional. It is also true that the federal system in Germany, which leaves university policy up to the states, threatens the international competitiveness of German universities. In practice, however, the differences between the existing *Habilitation* system and the junior professorship are marginal at best.

After completing a PhD and a postdoc, a *Habilitand* can join a research group at a German university. Many have no teaching duties and perform independent research, supervising undergraduate and graduate students, writing proposals for grants and publishing as much as they can. A few are well funded by the excellent Emmy Noether programme. Within six years, a

Habilitand has to summarize his or her research achievements: this summary, which is very similar if not identical to most tenure files, is subject to external review. Following an internal university commission, the faculty either grants or rejects the *Habilitation*; this is similar in many ways to a tenure procedure.

The German academic anachronism is not the *Habilitation* procedure itself but what follows. Whereas successful US applicants receive tenured positions at their universities, the German *Habilitand* has to apply for positions elsewhere and sometimes has to wait for years before achieving a tenured associate professorship (C3 status) or a full professorship at another university. In Germany, as elsewhere, the best young scientists will have multiple offers, some before they even finish their *Habilitation*. But for many others — too many — a long, unproductive and insecure phase follows.

Here the junior professorship could indeed help. Unfortunately, it does not, because, as noted in your News story, the German-style junior professorship is not linked to tenure track, forcing junior professors into the same fate as their *Habilitand* colleagues. Both are waiting and competing for rare openings at other universities. In the end, the differences between the two systems are too small to make much impact on the careers of young German scientists.

Thomas Carell

Department of Chemistry, University of Munich,
Butenandstrasse 5-13, D-81377 Munich, Germany

A duet on speciation

A synthesis of ideas about how species arise hits the right note.

Speciation

by Jerry A. Coyne & H. Allen Orr
Sinauer: 2004. 545 pp. \$89.95 (hbk); \$54.95, £35.99 (pbk)

Axel Meyer

Identifying and understanding the processes that lead to the origin of species is one of the fundamental problems of evolution. Solving it requires not only field studies, comparative analyses and lab work, but also a scholarly knowledge of previous work. *Speciation*, a joyous duet by two of the field's leading soloists, Jerry Coyne and Allen Orr, should help. This treasure trove of knowledge, which provides ample food for thought, succinctly outlines hypotheses and proposes research programmes to test them.

The literature on speciation spans more than a century and a half, and includes contributions from experimenters, innovative thinkers and theoreticians. A vast jungle of empirical data has accumulated, but every single tree must still be inspected carefully to develop a comprehensive picture of the entire forest. The empirical evidence must then be analysed rigorously to see how it fits with alternative models, to distinguish between competing hypotheses. Coyne and Orr have succeeded in this task.

But why, almost 150 years after the publication of Darwin's *On the Origin of Species*, is there still so much left to say and do? Surprisingly, Darwin did not directly address the question of the origin of species in his published writings. For him, the origin of adaptations to new ecological challenges within species and the origin of the species themselves seemed almost the same thing. Arguably, the processes that result in the divergence of lineages to form new species are different from the mechanisms of the longer, and often more pronounced, divergence that occurs afterwards. It is generally accepted that the processes differ in timescale: whether species arise rapidly (for example through polyploidization, as in some plants) or more gradually, the splitting of lineages is generally much faster than the subsequent divergence within a species. But are the processes through which divergence occurs during and after speciation fundamentally different? Coyne and Orr would probably say yes; others might disagree.

The study of speciation depends on the 'species problem' — knowing what constitutes a species. How objective are the criteria that we use to assign individuals and populations to units that we call species, and do they affect the study of the processes of



Going separate ways: divergence and speciation in the herring family.

speciation? Darwin largely ignored this issue as well and, at least in *On the Origin of Species*, seemed to treat species as if they are not even real in the philosophical sense. To him, species were convenient constructs — groupings of sets of individuals that resemble each other, but which are not distinct, being just artificially united fluctuating forms that overlap with other potential species. To overstate the dilemma, Darwin neither believed in species nor did he write about how new species arise. He left these really hard nuts for subsequent generations of evolutionary biologists to crack.

Ernst Mayr, the centenarian doyen of evolutionary biology, contends that the species problem can be reduced to a simple choice between species as realities of nature or as theoretical constructs of the human mind. There are many alternative definitions of species, each emphasizing some cause or biological feature of species, and/or basing its definition on a particular analytical or empirical approach or data set in an attempt to capture the essence of what makes a group of individuals a species. (Thankfully, these are discussed in the appendix of the book.) Some researchers have even advocated abolishing the notion of species altogether.

Can we even begin to answer questions about speciation without knowing what

species are? Having grown impatient over what they feel are pointless philosophical or (worse) semantic debates over the species problem, Coyne and Orr advocate the collection of more data on speciation and fewer debates on what species are. They conclude that it doesn't really matter to which particular species concept one subscribes.

I have painted a picture that is bleaker than it really is. Most biologists since Darwin have believed that species are real. The biological species concept (BSC) proposed by Mayr and Theodosius Dobzhansky is the most prominent and widely accepted one. It defines species as "groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups". Thus, new species arise when

reproductive isolation separates formerly interbreeding populations.

Like most researchers, Coyne and Orr feel that the study of speciation is helped, or even only becomes possible, by accepting the BSC and focusing on the study of reproductive isolation. They want to know where, how and when reproductive isolation comes about, and how it can be measured. However, some researchers believe that accepting the BSC stifles the study of speciation by limiting it to the study of the origin of reproductive isolation. The concept's dependence on reproductive isolation among sexually reproducing organisms, for example, means it cannot easily be applied to bacteria and organisms that reproduce clonally.

Geographic mechanisms of speciation, in which formerly interbreeding populations are prevented from doing so by geographic barriers, are generally thought to be the most common way for new species to arise, and this idea receives much empirical support. A non-adaptive by-product of isolation is that, over many generations, genomes or gene frequencies drift apart. This then prevents interbreeding and the homogenizing flow of genes from one species to the other when they come into contact again.

It was Mayr who laid the foundations for much of the current research on the origin

of species, in particular for the BSC and for most geography-based hypotheses of speciation. Coyne and Orr could be accused of being 'admayrers', but that's not the worst thing one can be called, in my opinion. Unlike Mayr in his early work, the authors admit, albeit with palpable hesitation, that speciation can also occur without geographic isolation, in overlapping or 'sympatric' areas. Sympatric speciation, which for decades had all the caché of a four-letter word, is increasingly seen to be responsible for the origin of at least a minority of species.

Coyne and Orr also give renewed credence to models of speciation that emphasize a role for sexual or natural selection, rather than viewing the origin of species as little more than a by-product of isolation. One currently popular model, ecological speciation, de-emphasizes geographical settings and reaffirms the importance of local adaptations and natural selection in bringing about speciation. Reproductive isolation may be brought about by ecological processes, such as habitat fragmentation or the uneven distribution of resources. In such cases, interbreeding is prevented between populations that are limited — by behavioural, physiological and morphological adaptations — to a very particular set of prey items, for example, or a particular tree species that they need to build their nests. The absence of the homogenizing effect of gene flow between individuals, not through geography but by the divergence of adapted and ecologically important traits, possibly even in the same environment, then leads to reproductive isolation. In this way, natural selection can play a prominent role in speciation.

The most recent work, largely done on fish models, includes modern genomic approaches, such as the analysis of quantitative trait loci. It concludes that reproductive isolation and speciation may be a by-product of ecological differences and disruptive selection on a surprisingly small number of phenotypic traits, which are controlled by an equally small number of underlying genes. A role for selection can also be inferred from molecular genetic data on hybrid incompatibilities from models such as the fruitfly. These have yielded a few candidate 'speciation genes', which also show signs of natural selection having been at work.

Performing this demanding duet in masterly harmony, Coyne and Orr present an authoritative treatise on one of the most long-running debates in evolutionary biology. *Speciation* is an impressively up-to-date and enlightening synthesis — and an entertaining read. It deserves to join Darwin's *On the Origin of Species*, and Mayr's *Systematics and the Origin of Species* on the bookshelf of anyone who is interested in evolution. ■

Axel Meyer is in the Department of Biology, University of Konstanz, 78457 Konstanz, Germany.



Grabbing a bite: the main role of teeth is to break down large items of food into smaller ones.

Something to chew on

Dental Functional Morphology: How Teeth Work

by Peter W. Lucas

Cambridge University Press: 2004. 372 pp. \$130, £75

Daniel E. Lieberman

Science has made substantial progress since Aristotle wrote (apparently without doing much research) that women have fewer teeth than men. The sheer volume of published research on teeth since may lead some to conclude that we have over-compensated for Aristotle's ignorance. Yet teeth merit all this attention because of their tremendous biological importance — not to mention the dreadful pain they can cause. Dental development and function are the focus of much clinical attention. And for evolutionary biologists, teeth are invaluable sources of information about taxonomy, phylogeny and many other aspects of animal biology.

There are already many excellent texts on dental function and development within the context of craniofacial development and clinical dentistry, as well as several good reviews of dental variation and evolution among vertebrates. But *Dental Functional Morphology* provides a fresh perspective on dental function. Peter Lucas's basic argument is that because the primary function of teeth is to reduce the size of food particles, dental morphology must be analysed in the context of how teeth fracture food, and how foods resist this. So the book reviews in detail many of the key mechanical properties of food, such as toughness and elasticity, which influence how teeth initially deform food items and generate cracks in them to break down large particles. Lucas then considers how variations in tooth size and shape

influence this. He also includes more general reviews of dental and oral anatomy, and provides an excellent summary of the processes of chewing and oral transport, viewed from the perspective of the mechanical properties of food, such as particle size and stickiness.

This food's-eye view leads to numerous insights and interesting ideas, such as Lucas's theory of fracture scaling. Bigger animals have bigger teeth, whose surface areas might be predicted to scale with body mass to the power of 0.67 (because tooth area increases to the power of two, and body mass increases to the power of three, yielding a scaling ratio of 2/3). Yet tooth surface area in mammals typically scales to the power of 0.61. Why this is so remains elusive, but Lucas argues that fracture mechanics plays a role.

The argument is complex, but boils down to the observation that once a crack is initiated in an object, little additional energy is needed to finish the job, regardless of its size. Bigger foods fracture at relatively low stress, which has several implications. One is that bigger animals (assuming that they chew bigger food) need relatively less muscle force (as quantified by muscle cross-sectional area), so this should only increase to the power of 0.5 relative to body mass, although this has yet to be tested. If tooth surface area does not increase relative to bite force, then tooth surface area should also scale to 0.5 relative to body mass. But teeth scale to the power of 0.61, so other factors must also influence tooth size, including other complex aspects of food mechanics also reviewed by Lucas. We can look forward to efforts to test this hypothesis and explore its implications.

Lucas does not consider in detail how dental function relates to tooth development and microstructure, or to the neuromuscular control of chewing. But readers interested in such topics as evolution, diet and ecology will enjoy his many other ideas about how vertebrate teeth work. The final chapter focuses mainly on mammals, creatively integrating

biomechanics, anatomy, ecology and taxonomy in order to reconsider the evolution of dental adaptations for generating fractures in food for animals that eat insects, grasses, leaves, fruit and other animals. Primates, especially humans, get special attention. In general, we chew our food like other mammals, but the invention of cooking and other forms of food processing have drastically decreased the particle size and toughness of the food we eat. Palaeoanthropologists are still arguing about when cooking first evolved, but Lucas provides new reasons to suggest that the first species of the genus *Homo*, which had small teeth, was the first true chef of the animal world. Lucas calculates that your molars can be between 56% and 82% smaller to eat a cooked potato rather than a raw one, depending on whether you eat the skin and whether you roast or boil it.

Teeth often appear messy, confusing and dull to non-specialists, but Lucas succeeds in conveying his enthusiasm for the challenges of learning about the biology and ecology of organisms from such a small and humble organ. Although the book contains plenty of mechanics, the equations are presented clearly and well explained.

Lucas has filled the text with fascinating observations, humorous asides and wonderfully detailed footnotes. One particularly fun bonus is a flick-art animation running on the bottom corner of the book's pages that depicts the evolutionary transformation of a primitive single-cusped tooth into a human molar. Flicking through this cartoon will give your copy a thumb-worn look that Aristotle might have envied.

Daniel E. Lieberman is in the Department of Anthropology, Harvard University, 11 Divinity Avenue, Cambridge, Massachusetts 02138, USA.

The geography of life

Foundations of Biogeography: Classic Papers with Commentaries

edited by Mark V. Lomolino, Dov F. Sax & James H. Brown

University of Chicago Press: 2004. 1,291 pp. \$135, £94.50 (hbk); \$45, £31.50 (pbk)

Mark Williamson

Collections of papers are useful both to undergraduates and their lecturers. Here is a set of 72 pieces on biogeography, 30 of which are excerpts from books. They were chosen and edited by a committee of 19 biologists, 13 of whom have contributed to the commentaries that precede the eight sections. Except for the first section, on early (mostly nineteenth century) classics, these are arranged by topic, which is helpful. In much the same way that a camel is a horse designed by a committee, the selection of topics has some strange bumps and depressions, but the resulting animal is nevertheless useful in the appropriate circumstances.

The editors suggest that biogeography is a recent discipline and that nine of the authors would not have called themselves biogeographers. But biogeography is a nineteenth-century term. Darwin wrote of the geographical distribution of organic beings, and Alfred Russel Wallace wrote of geographical zoology. And in the twentieth century, Robert MacArthur and Edward O. Wilson are in print saying "we both call ourselves biogeographers" and are "unable to see any real distinction between biogeography and ecology".

Biogeography might appear to be an

interdisciplinary subject between biology and geography. It is certainly taught in both sets of departments, which once caused me a little difficulty as an external examiner because students had been taught essentially the same course twice. But biogeography is a branch of biology, of population and community ecology, and covers genetical and evolutionary topics as well as pure ecological ones. It is certainly an important branch, although the editors exaggerate its importance to Darwin and Wallace in discovering natural selection. It is nevertheless a branch with unusually fuzzy edges, differing from the related bits of biology largely by an emphasis on maps and geology.

With so many editors involved, the standards of the commentaries and the editing are both a bit variable, though it is mildly amusing to be told about "Louis Carroll's *Through the Looking Glass*" (two errors there) or that the muskrat "was introduced in 1905 into what was then Czechoslovakia". More seriously, in my view, the editorial board (who are all biologists) would have been well advised to seek the opinions of historians of science on the works from before 1950. Using facsimiles that have been standardized to a fixed page size means that some of the text is hard to read and some half-tones would have been better omitted. The cross-referencing is a bit weak, too. Some of the commentaries are excellent.

If your class reading list calls for something on Linnaeus, Buffon, de Candolle, von Humboldt and Hooker at one end of the time span, along with MacArthur and Wilson's theory and developments in the 1970s, this 2-kg set will be very useful.

Mark Williamson is in the Department of Biology, University of York, York YO10 5DD, UK.

Science in culture

Awash with art

Zeger Reyers saturates the senses with his installation *Aqua Boogie*.

Colin Martin

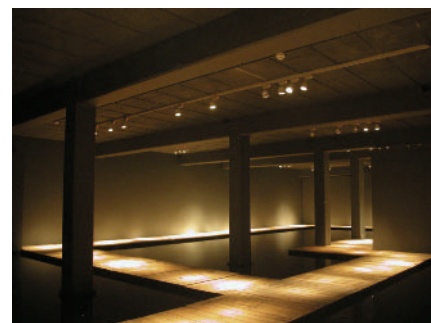
The installation by Dutch artist Zeger Reyers in the basement of the GEM Museum of Contemporary Art in the Hague, the Netherlands, acknowledges the prior claim of water on the museum's site. The land was drained by previous generations.

Reyers' work frequently creates new biotopes, which display the vagaries of biological processes. At the 2003 Havana Biennial he released 200 white laboratory mice onto a 15-metre-long white tabletop covered with 5,000 kg of white porcelain. Their colouring made them difficult to spot, creating an installation that referenced protective adaptation in nature. Other previous installations covered interior spaces with living fungi.

To reach *Aqua Boogie*, visitors descend a staircase that has a view across a large pond

outside, conveying a feeling of being below water level. They enter an irregular, 220-square-metre space, flooded with dark, non-reflecting water and criss-crossed with pine duckboards (right). As well as allowing viewers to explore the aqueous biotope more closely, this grid recalls *Victory Boogie Woogie*, a painting by Piet Mondrian, as the installation's title suggests.

The biotope affects many of the senses. "It has the thin fragrance of living water, like you smell when entering a canal," says Reyers. "The water also dampens sounds, making it a quiet, serene place." Beneath the surface, 35 leather carp, selected by Reyers because their dark colouring makes them practically invisible, go about unseen business. Like all Reyers' work, *Aqua Boogie* is accessible, but offers multiple levels of interpretation. It can disconcert viewers by challenging their perceptions of



a familiar environment, experienced out of context.

On two occasions during the installation, which can be seen until 7 November, Reyers will recreate a 2001 work, *Mussel Chair*. A Parisian pavement café chair, encrusted with mussels after having been submerged in the Eastern Scheldt estuary for two years, will be brought to the museum, cleaned up and steamed. The cooked shellfish will be offered to visitors to eat, adding taste to the other sensory experiences offered by *Aqua Boogie*.

Colin Martin is a writer based in London.

Mending and malignancy

Carcinogenesis and tissue repair: how might chronic tissue injury lead to the development of cancer?

Philip A. Beachy, Sunil S. Karhadkar and David M. Berman

The familiar sensation of heartburn — whether from greasy food or too much coffee — is caused by acid reflux from the stomach into the oesophagus and is probably harmless on an occasional basis. On the other hand, persistent heartburn can indicate lasting damage to the lining of the oesophagus, resulting in changes in the cellular architecture of the epithelium and an increased incidence of the lethal cancer oesophageal adenocarcinoma. Increased cancer risk is also associated with other types of chronic tissue injury, including those caused by toxins (such as alcohol, cigarette smoke and solvents), chronic infection of *Helicobacter pylori* and other pathogens, and inflammatory conditions, such as sclerosing cholangitis and inflammatory bowel disease. Genetic mutations are thought to be at the root of cancer, but although some of these injurious agents are mutagens, others are not. How, then, does injury contribute to cancer risk?

To answer this question, we consider tissue response to injury. When faced with injuries that might compromise their function as barriers, the epithelial linings of organs (such as the airway and digestive tract) respond rapidly to restore their integrity. This repair programme includes movement of cells to patch defects, as well as rapid production of cells that can differentiate to form the specialized cells needed to repopulate the injured epithelium. The ultimate source of these differentiated cells within the injured tissue is stem cells.

Stem cells are also increasingly being viewed as playing a role in cancer. Their capacity for self-renewal and unlimited replication make them appealing candidates as the cell of origin for cancer. In addition, stem cells are long-lived, providing ample opportunity to accumulate the mutations that could cause an increase in the rate of cell proliferation and produce clinically significant cancers. There is some evidence to support these ideas. In acute myelogenous leukaemia, the small fraction of leukaemic cells capable of propagating the cancer are undifferentiated, with a combination of surface markers similar to those of blood stem cells found in bone marrow. Elsewhere in the body, the discovery of multipotent progenitor cells that have the capacity for self-renewal opens the possibility that other cancer types also could arise from their respective tissue stem cells.



Adding insult to injury: can tissue damage lead to cancer?

Another common trait of both cancer and the activated state of repair is the activation of signalling pathways best known for their roles in embryonic growth and patterning. These include the Hedgehog (Hh) and Wnt signalling pathways, which also act post-embryonically to stimulate stem cell self-renewal. The Hh and Wnt pathways are also chronically activated in, and required for, the growth of many cancers, including carcinomas that arise in the oesophagus, stomach, lung, pancreas, biliary tract, colon and prostate — which together account for as many as one-third of cancer deaths.

Thus, activation of the Hh and Wnt signalling pathways not only promotes stem cell self-renewal during tissue regeneration, but also the growth of cancers, which may arise from stem cells. Could it be that cancer represents the continuous operation of an unregulated state of tissue repair associated with chronic activation of pathways such as Hh and Wnt? At the cellular level, the simplest form of this idea would be that a cancer is initiated when a stem cell activated by a regenerative programme fails to return to the resting state once the repair is completed. The events that lock this cell in an activated state and produce a clinically significant cancer could include multiple genetic and/or epigenetic changes, as well as influences from surrounding tissues. The motile behaviour of cells in this activated state, which helps to patch injury-induced epithelial defects, may help to account for the highly invasive nature of some of these epithelial cancers (although it is not yet clear whether epithelial stem cells themselves migrate).

How does this hypothesis link chronic injury to cancer, and how can differences in carcinogenic efficiency of injurious agents

be explained? The overall rate of cancer formation would be predicted to depend on the number of activated stem cells that can initiate cancer growth.

Any repeated stimulus leading to a chronic increase in stem-cell activation over time should therefore result in a higher overall frequency of cancer formation. This stimulus could come from exogenous toxins — such as alcohol, which causes liver injury and an increased rate of liver cancer — or from endogenous factors, such as gastric acid, which can cause oesophageal injury and an increased rate of oesophageal cancer.

But it could also come from normal

physiological events, such as injury to the ovarian epithelium during ovulation, the suppression of which by multiple pregnancies or oral contraceptive use is associated with reduced rates of ovarian cancer. In potent carcinogens such as cigarette smoke, the presence of mutagens in addition to injurious agents could accelerate the oncogenic process by expanding activated stem-cell pools while simultaneously inducing genetic changes in these cells. A combination of expanded stem-cell pools and mutagenesis might also account for the efficiency of carcinogenesis in tumour initiator/promoter models, which combine mutagenic agents such as ethylnitrosourea with agents that induce cell proliferation, such as phorbol esters. A better understanding of the response to epithelial injury, in particular the signals that regulate the activation of tissue stem cells, might lead to useful strategies for cancer prevention and therapy.

The authors are affiliated with the Department of Molecular Biology and Genetics and the Howard Hughes Medical Institute (P.A.B. and S.S.K.), and with the Departments of Oncology (P.A.B. and D.M.B.), Pathology (S.S.K. and D.M.B.) and Urology (D.M.B.) at The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.

FURTHER READING

- Taipale, J. & Beachy, P. A. *Nature* **411**, 349–354 (2001).
- Reya, T. *et al.* *Nature* **414**, 105–111 (2001).
- Coussens, L. M. & Werb, Z. *Nature* **420**, 860–867 (2002).
- Al-Hajj *et al.* *Proc. Natl Acad. Sci. USA* **100**, 3983–3988 (2003).
- Watkins, D. N. *et al.* *Nature* **422**, 313–317 (2003).
- Karhadkar, S. S. *et al.* *Nature* advanced online publication, 12 September 2004 (doi:10.1038/nature02962).

Ethics and amphibians

Robert M. May

A statistical study shows convincingly that a technique for marking frogs in ecological field experiments compromises the results. Present practices need a rethink — and not only for practical reasons.

It was 25 years ago, at an ecology seminar at Princeton University, that I first learned of the standard method for 'marking' individual newts or other amphibians (Fig. 1) by clipping their toes. In this way, each individual can be identified by the unique combination of digits removed. I remember being impressed by the elegance of the experiments concerned — but even more impressed by the casual barbarity of the toe clipping.

Seeking to avoid a larger ethical minefield, I asked whether such removal of digits would affect survival, particularly in more heavily clipped individuals, thus compromising the conclusions. My question was swept aside as silly (the sort of thing you might expect a theoretician to ask). But it now appears to have been answered. Writing in the *Journal of Applied Ecology*, McCarthy and Parris¹ find that "toe clipping reduces the return rate [recapture of marked individuals] by 4–11% for each toe removed after the first, assuming the effect is the same for all toes".

The intervening years have, indeed, seen studies of the possible adverse effects of toe clipping of amphibians, including inflammation and infection of feet and limbs. Some of these studies indicate lower rates of return of marked individuals, others have found no such effects; Williamson and Bull² give a good review.

In an earlier paper, Parris and McCarthy³ suggested that these apparent inconsistencies arise from a lack of statistical power (sample sizes too small for statistically significant conclusions to be drawn) in some previous work, rather than from the absence of real effects. They estimated that return rates decline by 6–18% for each toe removed after the first. In this earlier paper³, however, they could not provide meaningful confidence intervals or a measure of how the effects would change with the number of toes removed.

The new paper¹ gives a Bayesian reanalysis of data drawn from four published studies on the return rate of toe-clipped frogs: 1,333 individuals of *Crinia signifera* (a small ground-dwelling Australian frog) with up to seven toes removed²; 306 individuals of the same species with between two and four toes removed⁴; 733 individuals of *Bufo fowleri* (a large ground-dwelling frog from eastern United States) with up to eight toes removed⁵; and 1,307 individuals of *Hyla labialis* (a medium-sized tree frog from the



Figure 1 Toe show. Frogs are often the subject of studies involving toe clipping.

Colombian Andes) with up to seven toe-disks removed⁶. The authors' Bayesian model, employing a freely available statistical program⁷, helps "define the relationship between the number of frogs that are recaptured and the effect of toe-clipping as a function of the number of toes removed"¹.

The estimated decline in return rate of 4–11% for each toe clipped assumed that the adverse effects are independent of the total number of toes clipped. McCarthy and Parris also provide a more refined analysis, allowing effects of toe removal to change linearly with the number removed. This indicated, for example, that the removal of a second toe reduced return rate by 3.5% (with a 95% "credibility interval" of 0–7%), whereas return rate was reduced by 30% (95% "credibility interval" of 20–39%) on removal of an eighth toe. Overall, the cumulative effect of toe clipping appears to be such that the return rate for frogs with two toes removed was around 96% of those with one toe removed, declining to 28% for the return of frogs with eight toes gone.

The authors justifiably conclude that these effects should be more explicitly recognized in future studies that use toe clipping to mark individuals in ecological studies. They also tellingly add that "our study has important implications for the ethical treatment of animals, the continued use of clipping to mark species of conservation concern, and the removal of multiple toes from an individual frog or toad". It

seems to me that all these conclusions apply to all amphibians.

More generally, I see McCarthy and Parris's paper as a notable addition to a growing literature that raises both practical and larger ethical questions about time-honoured procedures in some ecological field studies. There are obvious parallels with the recent study of long-term effects of flipper tags on penguins, by Gauthier-Clerc *et al.*⁸. This work attracted considerable media attention with its finding, after five years' work on king penguins implanted with electronic tags (some also with flipper bands and others not), that "banding results in later arrival at the colony for courtship in some years, lower breeding probability and lower

chick production". Gauthier-Clerc *et al.* also found that unbanded king penguin chicks had roughly twice the survival rate after 2–3 years of those encumbered with flipper bands.

Scientific and medical understanding gained by the use of non-human animals in laboratory studies is widely recognized as producing great benefits, primarily for humans but also for other animals themselves. At the same time, such research is properly strictly regulated. And I for one welcome the attention increasingly given to the rights of non-human animals by philosophers such as Peter Singer. But field studies of the ecology and behaviour of non-human animals can also raise difficult questions of costs and benefits. As a sixth wave of mass extinction looms, conservation biologists desperately need the knowledge that comes from such field studies. As the work described here clearly shows, however, there are good reasons why we need to think more carefully about some present practices. ■

Robert M. May is in the Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.
e-mail: robert.may@zoology.oxford.ac.uk

1. McCarthy, M. A. & Parris, K. M. *J. Appl. Ecol.* **41**, 780–786 (2004).
2. Williamson, I. & Bull, C. M. *Wildl. Res.* **23**, 249–266 (1996).
3. Parris, K. M. & McCarthy, M. A. *Amphibia-Reptilia* **22**, 275–289 (2001).
4. Lemckert, F. *Amphibia-Reptilia* **17**, 287–290 (1996).
5. Clarke, R. D. *Copeia* **1972**, 182–185 (1972).
6. Lüddecke, H. & Amézquita, A. *Copeia* **1999**, 824–830 (1999).
7. Spiegelhalter, D. J. *et al.* *WinBUGS User Manual*, Version 1.4 (MRC Biostatistics Unit, Cambridge, UK, 2003).
8. Gauthier-Clerc, M. *et al.* *Proc. R. Soc. Lond. B* doi:10.1098/rsbl.2004.0201 (2004).

Break-up breakdown

Tim Reddish

Molecules ionize and fragment when subjected to energetic radiation. The behaviour of a simple molecule, deuterium, can now be tracked through this process in greater detail than ever before.

It is said that when physicists want to know what is going on inside a microscopic object, they just blow it up and see what happens. This reductionist approach has certainly proved useful in particle physics and in studying the structure of atoms, molecules and nuclei, using a variety of projectiles. The explosive experiments reported by Weber *et al.*¹ on page 437 of this issue demonstrate that even simple molecular hydrogen is full of surprises. Their study addresses the fundamental question of what exactly happens to an atom or molecule when it absorbs sufficient energy to fragment completely. Put another way, what governs the motions of the free charged particles in a so-called Coulomb explosion?

If a molecule absorbs enough energy, bonds can be broken and ions can form, allowing neutral and charged constituents to react with neighbouring molecules. These processes occur everywhere, from car engines to industrial chemical plants. Molecular ionization and fragmentation also occur naturally in Earth's upper atmosphere, driven by the Sun's energy, and are the mechanisms responsible for radiation damage to our bodies. In general, the more energy that is absorbed — either from light or through a collision with another energetic particle — the greater the number of possible reaction pathways and the greater the 'damage' done to the molecule itself and its nearest neighbours.

The removal of two electrons from helium, the simplest multi-electron atom, leaves behind a completely bare nucleus — an α -particle. Thus, the double-ionization process produces three charged particles: the positively charged nucleus, and two electrons that escape under the influence of the long-range Coulomb force. In the case of molecular hydrogen, once the electrons have been ejected, the 'glue' that held the molecule together has gone and the protons explode apart in opposite directions, because like charges repel. One of the initially surprising results, for both helium and hydrogen, was that if the two escaping electrons have equal energies, they do not leave in opposite directions, as might be expected. Quantum mechanics, as well as the Coulomb force, determines the outcome.

It should be emphasized that, even when it is energetically possible, the double-ionization process in helium and hydrogen is still about 30 times less likely to occur than the

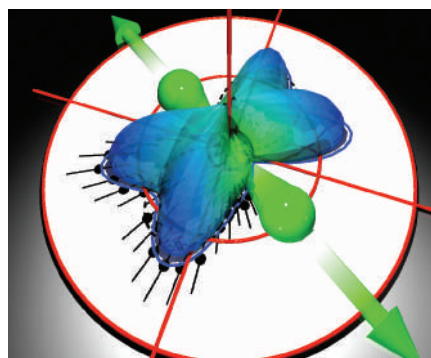


Figure 1 A three-dimensional 'photograph'¹ of the double ionization of deuterium molecules, initiated by the absorption of photons. The surface represents a model fit to a sample of Weber and colleagues' data¹, showing the angular distribution of one ejected electron in the plane containing the molecular and light polarization axes; the other escaping electron (of the same energy) is emitted directly upwards, out of the plane. The direction of the molecular axis is given by the exploding nuclei (green), whose energies have been selected such that the vibrating nuclei are as close together as possible at the instant of photoabsorption.

ejection of a single electron. This is because the process for the simultaneous emission of two electrons is dominated by interactions between the electrons themselves — they must be treated as a correlated pair and not as independent particles — and this improbability makes the experiments much more difficult. Ultrafast timing techniques are used to ensure that the pairs of electrons detected really do come from the same atom or molecule. Even so, as the two electrons can be emitted in any spatial direction, efficient methods of particle detection are needed that cover as large a solid angle as possible. Completely new instruments have been invented to enable the fragment products to be detected and to measure both their energies and their angles. Some groups have developed doughnut-shaped 'toroidal' analysers to view the process in one plane^{2–4}, while others have produced an elegant three-dimensional detection system^{5–7}. All of these experiments have benefited from technological advances at synchrotron sources, as ideally they require extreme-ultraviolet radiation with 100% linear polarization to initiate the Coulomb explosion.

The new study by Weber *et al.*¹ builds on earlier experimental studies, by this group^{5,6}

and others^{2–4,8–10}. Their three-dimensional momentum-imaging approach has reached new heights of sophistication, giving a view in unprecedented detail of the molecular double-ionization process for the hydrogen isotope deuterium. The experimental set-up is surprisingly simple in principle. In the central interaction region, deuterium molecules are ionized by a beam of synchrotron light. A controlled, uniform electric field surrounds the interaction region and accelerates the two positive ions in one direction and the electrons in the opposite direction. After the charged particles have left the electric field, they drift onto position-sensitive detectors. The final ion energies (about 9.5 electronvolts) are governed by the repulsive Coulomb force between the two deuterium nuclei and depend on their initial separation in the neutral molecule. The electrons, however, can have any amount of energy — as long as energy is conserved overall.

A uniform magnetic field is imposed along the symmetry axis of the apparatus to constrain the trajectories of fast electrons and to ensure that they are able to reach the detector. This creates complicated spiral trajectories for the electrons but does not significantly perturb the motion of the much heavier ions. For each charged particle, the position of impact on the detector and the overall 'time of flight' are measured, and from them the initial momentum of all four particles can be deduced directly. Even more excitingly, in the case of hydrogen, the ions leave in opposite directions so quickly (compared with their rotational motion) that even the alignment of the molecular axis can be determined with respect to the polarization direction of the incident light. Thus, this gives a three-dimensional 'photograph' of the Coulomb explosion.

Weber *et al.*¹ have, for a specified direction of one electron, determined the spatial distribution of the other escaping electron as a function of the alignment between the molecular and polarization axes in the double ionization of deuterium (Fig. 1). From their data, it is clear that the degree of alignment of the molecular axis affects the escape directions of the electrons and that those directions do not depend solely on the polarization of the synchrotron light. This is no surprise, but the fact that the effect can now be measured is truly impressive. As if that were not enough, by carefully selecting the energy of the ions, the electron emission patterns can be probed as a function of the initial internuclear separation of the ion cores, just before the double ionization happens. This sensitive three-dimensional molecular camera captures images of the Coulomb explosion at various bond lengths, as the molecular bond stretches and contracts through its own vibrational motion.

In the past, theories have generally had to integrate over molecular orientation and

internuclear separation — smearing out interesting physics — just to be able to compare with experiment. Those days are now apparently over. These remarkable experimental results create a timely theoretical challenge: to describe the observations accurately and predict new areas of interest. The next obvious development is to study a molecule comprising one hydrogen atom and one deuterium atom — still a simple molecule, but with its symmetry broken. In this case, the orientation of the molecular axis, not just its alignment, can be determined — that is, which end is 'up', not just which way it is pointing; chiral effects are also anticipated with circularly, instead of linearly, polarized light. In terms of the 'big picture', these studies will provide remarkable physical insight

into multi-electron processes whose effects are even more significant in larger atoms and molecules.

Tim Reddish is in the Department of Physics, University of Windsor, Windsor, Ontario N9B 3P4, Canada.
e-mail: reddish@uwindsor.ca

1. Weber, T. *et al.* *Nature* **431**, 437–440 (2004).
2. Reddish, T. J. *et al.* *Phys. Rev. Lett.* **79**, 2438–2441 (1997).
3. Wightman, J. P., Cvejanović, S. & Reddish, T. J. *J. Phys. B* **31**, 1753–1764 (1998).
4. Secombe, D. P. *et al.* *J. Phys. B* **35**, 3767–3780 (2002).
5. Dörner, R. *et al.* *Phys. Rev. Lett.* **81**, 5776–5779 (1998).
6. Weber, T. *et al.* *Phys. Rev. Lett.* **92**, 163001 (2004).
7. Dörner, R. *et al.* *Phys. Rep.* **330**, 95–192 (2000).
8. Kossmann, H., Schwarzkopf, O., Kämmerling, B. & Schmidt, V. *Phys. Rev. Lett.* **63**, 2040–2043 (1989).
9. Dujardin, G., Besnard, M. J., Hellner, L. & Malinovich, Y. *Phys. Rev. A* **35**, 5012–5019 (1987).
10. Scherer, N., Lörch, H. & Schmidt, V. *J. Phys. B* **31**, L817–L822 (1998).

Cancer

An inflammatory link

Fran Balkwill and Lisa M. Coussens

The NF- κ B protein is a key player in inflammation. It now seems that it might also activate signalling pathways, in both cancer cells and tumour-associated inflammatory cells, that promote malignancy.

Inflammation is central to our fight against pathogens, but if it is not ordered and timely, the resulting chronic inflammation can contribute to diseases such as arthritis, heart attacks and Alzheimer's disease. A functional link between chronic inflammation and cancer has also long been suspected^{1,2}: population-based studies show that susceptibility to cancer increases when tissues are chronically inflamed; and long-term use of non-steroidal anti-inflammatory drugs reduces the risk of several cancers³. Moreover, most solid tumours contain many non-malignant cells, including immune cells and blood-vessel cells, that are important in inflammation. But the crucial molecular pathways that permit communication between abnormally growing cancer cells and these inflammatory cells remain unknown. A complex network of pro-inflammatory mediators is probably involved, because deletion of certain key molecules can reduce cancer susceptibility in mice^{1,2,4}. Two mouse models of inflammation-associated cancer now implicate the gene-transcription factor NF- κ B and the inflammatory mediator known as tumour-necrosis factor- α (TNF- α) in cancer progression^{5,6}.

Using a mouse model of inflammatory hepatitis that predisposes mice to liver cancers, Pikarsky *et al.*⁵, writing on page 461 of this issue, present evidence that the survival of hepatocytes — liver cells — and their progression to malignancy are regulated by NF- κ B. (NF- κ B is an important transcription factor that controls cell survival by regulating programmed cell death, proliferation

and growth arrest.) Moreover, Pikarsky *et al.* find that the activation state of NF- κ B, and its localization in the cell, can be controlled by TNF- α produced by neighbouring inflammatory cells (known collectively as stromal

cells). Greten *et al.*⁶, reporting in *Cell*, come to a similar conclusion by studying a mouse colitis-associated cancer model. Their work does not directly implicate TNF- α , but instead found enhanced production of several pro-inflammatory mediators ('cytokines'), including TNF- α , in the tumour microenvironment during the development of cancer.

An important feature of both studies is that NF- κ B activation was selectively ablated in different cell compartments in developing tumour masses, and at different stages of cancer development. These approaches offer new insight into the differential regulation of pre-malignant and malignant states by inflammation and NF- κ B in distinct cellular compartments.

In Pikarsky and colleagues' inflammation-associated model of liver cancer⁵, TNF- α , produced by stromal cells, activated NF- κ B in adjacent hepatocytes that were undergoing 'transformation' into malignant cells. Selective deletion of NF- κ B in hepatocytes, or inhibition of TNF- α produced by stromal cells, induced programmed cell death of transformed hepatocytes, and subsequently reduced the incidence of liver tumours. The authors also found that activation of NF- κ B was not important in the early stages of tumour development, but was crucial for malignant conversion. Thus, preventing NF- κ B activation in hepatocytes after seven months of chronic liver inflammation was sufficient to inhibit the

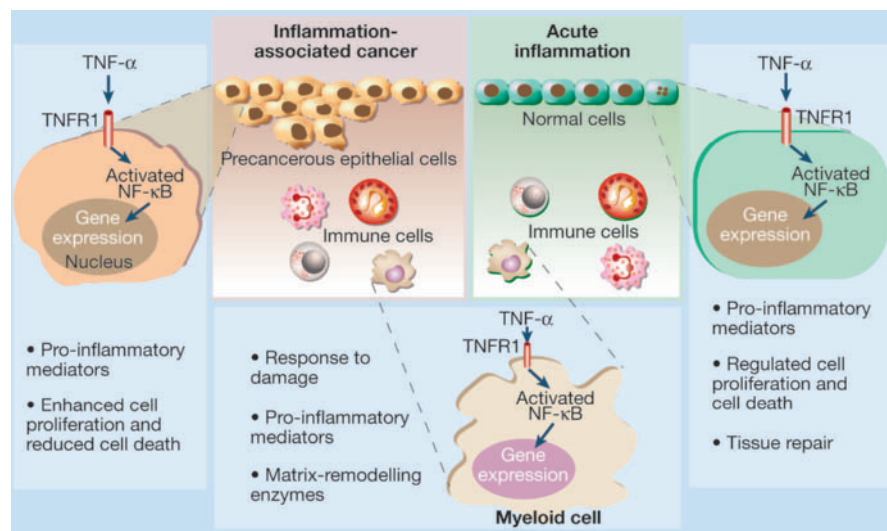


Figure 1 Opposing effects of the NF- κ B protein in normal tissues and in cancer. Tumour-necrosis factor- α (TNF- α) acts through its receptor, TNFR1, to activate the gene-transcription factor NF- κ B. During acute inflammation in 'normal' epithelial cells (right), NF- κ B activation leads to increased expression of genes that encode pro-inflammatory mediators called cytokines, and activates genes that regulate the balance between cell proliferation and cell death. In inflammatory immune cells (myeloid cells; bottom), NF- κ B activation can also regulate cell death, but more importantly regulates short-term expression of pro-inflammatory mediators to repair the tissue damage. Ablation of NF- κ B in inflammatory cells impairs the expression of pro-inflammatory cytokines. Precancerous epithelial cells (left) use NF- κ B to enhance their survival, and their propensity to become malignant cells, by augmenting their expression of pro-inflammatory and cell-survival genes while inhibiting the death-promoting machinery. If NF- κ B activation is disabled in these epithelial cells, while pro-inflammatory gene programmes are maintained, cell death is favoured and tumour progression is reduced.

development of liver cancers. So TNF- α , produced by surrounding stromal cells, activated an NF- κ B-dependent pathway that inhibited cell death during the period when precancerous hepatocytes developed into cancers.

In the model described by Greten *et al.*⁶, selective deletion of IKK- β — a key intermediary of NF- κ B — in cells of the intestinal lining did not decrease intestinal inflammation, but did reduce the subsequent development of intestinal tumours. As in the liver-cancer model, preventing NF- κ B activation led to the death of more of the intestinal cells that would otherwise have given rise to cancer. However, Greten *et al.* went one stage further by generating mice in which IKK- β was selectively deleted in inflammatory cells that infiltrate pre-malignant and malignant tumours. In these mice, the levels of messenger RNAs encoding several pro-inflammatory cytokines decreased, as did subsequent tumour development.

Cancers develop in phases. First there is thought to be an initiating event — genetic damage that renders an epithelial cell (such as a hepatocyte or a cell lining the intestine) capable of forming a cancer. Initiated precancerous cells then multiply during a promotion phase, when further genetic damage is incurred. And, in the case of inflammation-induced cancer, non-genetic stimuli also encourage the survival and proliferation of initiated cells.

The data obtained in these two cancer models^{5,6} suggest that the NF- κ B pathway does not affect initiation but has dual actions in tumour promotion: first by preventing the death of cells with malignant potential, and second by stimulating the production of pro-inflammatory cytokines in inflammatory cells in the tumour mass. The pro-inflammatory cytokines then signal to initiated or otherwise 'damaged' cells to promote their survival and proliferation (Fig. 1). This model is consistent with other studies^{7–9} and with observed correlations between the numbers of inflammatory cells, levels of cytokines and tumour severity and prognosis in both mice and humans⁴.

However, in some human and mouse cancers, the transformed cells themselves can also contribute to the overall levels of soluble pro-inflammatory cytokines. For example, TNF- α produced by skin cells (keratinocytes) is a tumour promoter in experimental skin cancer, where it produces a cascade of cytokines and of enzymes that degrade the extracellular matrix during the early stages of tumour promotion⁷. (The transcription factor implicated here, however, is AP-1 rather than NF- κ B¹⁰.) Human keratinocytes also produce TNF- α in response to ultraviolet irradiation¹¹. Other studies show that TNF- α produced by hepatocytes (rather than by neighbouring inflammatory cells) acts as a hepatocyte growth factor, and that mice lacking TNF receptors are resistant to chemically

induced liver carcinogenesis¹². Moreover, as Pikarsky *et al.* and Greten *et al.* state, activation of NF- κ B occurs in many malignant cells as a result of genetic mutation rather than in response to signals from surrounding cells. So pro-inflammatory cytokines contribute to tumour promotion not only by signalling from tumour-associated inflammatory cells to precancerous cells, but also through production in the precancerous cells themselves, especially at early stages.

Should future anticancer strategies focus on regulating NF- κ B activation or the availability of TNF- α ? It is important to note that all organs are endowed with unique cell-death pathways, as well as with damage-response pathways that typically involve short-term activation of innate immune cells. In the skin, for example, keratinocytes die through 'terminal differentiation', and inhibiting NF- κ B in initiated keratinocytes actually promotes one type of cancer by reducing terminal differentiation¹³. Conversely, blocking TNF- α attenuates chemically induced skin-tumour formation¹⁴. The idea of therapeutically regulating TNF- α , however, must also be thoroughly investigated, as this cytokine, too, has opposing activities that depend on the cell type and environment¹⁵. Phase I clinical trials of TNF- α antagonists are currently under way in patients with advanced cancer, and might help researchers to understand these issues^{14,15}. The new results^{5,6} also suggest that any anti-inflammatory therapy would be most effective during the early stages of cancer development.

Tissues have a complicated architecture,

and their cells respond to damage and die through inflammation-dependent and -independent mechanisms. So it is no wonder that inflammatory mediators such as NF- κ B and TNF- α have many different effects, depending on the context in which they are called into play. The studies by Pikarsky *et al.*⁵ and Greten *et al.*⁶ articulate these complexities, and show that, in devising anticancer strategies, we must consider the context in which tumour cells thrive, as well as the cells themselves. ■

Fran Balkwill is in the Cancer Research UK Translational Oncology Laboratory, Barts and The London Queen Mary's Medical School, Charterhouse Square, London EC1M 6BQ, UK. e-mail: frances.balkwill@cancer.org.uk

Lisa M. Coussens is at the Cancer Research Institute, Department of Pathology and Comprehensive Cancer Centre, University of California at San Francisco, 2340 Sutter Street, San Francisco, California 94143, USA.

e-mail: coussens@cc.ucsf.edu

1. Balkwill, F. & Mantovani, A. *Lancet* **357**, 539–545 (2001).
2. Coussens, L. M. & Werb, Z. *Nature* **420**, 860–867 (2002).
3. Gupta, R. A. & Dubois, R. N. *Nature Rev. Cancer* **1**, 11–21 (2001).
4. Balkwill, F. *Nature Rev. Cancer* **4**, 540–550 (2004).
5. Pikarsky, E. *et al.* *Nature* **431**, 461–466 (2004).
6. Greten, F. R. *et al.* *Cell* **118**, 285–296 (2004).
7. Moore, R. J. *et al.* *Nature Med.* **5**, 828–831 (1999).
8. Voronov, E. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 2645–2650 (2003).
9. Enzler, T. *et al.* *J. Exp. Med.* **197**, 1213–1219 (2003).
10. Arnott, C. H. *et al.* *Oncogene* **21**, 4728–4738 (2003).
11. Kock, A. T. *et al.* *J. Exp. Med.* **172**, 1609–1614 (1990).
12. Knight, B. *et al.* *J. Exp. Med.* **192**, 1809–1818 (2000).
13. Seitz, C., Lin, Q., Deng, H. & Khavari, P. *Proc. Natl Acad. Sci. USA* **3**, 2307–2312 (1998).
14. Scott, K. A. *et al.* *Mol. Cancer Ther.* **2**, 445–451 (2003).
15. Szlosarek, P. & Balkwill, F. *Lancet Oncol.* **4**, 565–573 (2003).

Global change

Carbon conundrum on the tundra

Wendy M. Loya and Paul Grogan

Vast amounts of carbon are locked into soils at northern high latitudes. The vexed question of how these ecosystems will respond to global warming is addressed by a long-term experiment in the Arctic.

On page 440 of this issue¹, Mack and colleagues describe how they dug deep into the carbon balance of an arctic tundra ecosystem, and came up with some surprising results. Their research reveals that tundra plants and soils respond in opposing ways to long-term nutrient fertilization. In their experiment, intended to simulate increased nutrient availability under warmer temperatures, plants grew better and stored more carbon, but valuable soil carbon was lost. In the context of global warming, the main implication of their findings is that the losses of deep-soil carbon that they observe could mean even greater increases in carbon dioxide concentrations in the atmosphere.

The carbon balance of terrestrial ecosystems is the difference between carbon

storage in plants and soils, and carbon losses resulting from decomposition of these tissues. In the cold, wet climate of high latitudes, decomposition proceeds slowly, and carbon accumulates in thick layers of organic matter on top of mineral soils. This net carbon storage in tundra and boreal forests accounts for an estimated one-third of the global soil-carbon pool², and is equivalent to two-thirds of the carbon presently found in the atmosphere. Hence the worry that changes in climate resulting in a warmer, drier environment might alter this balance through accelerated decomposition of soil organic matter. This could result in net carbon losses, positive feedback into increases in atmospheric concentrations of CO₂, and accentuated climate change. On the

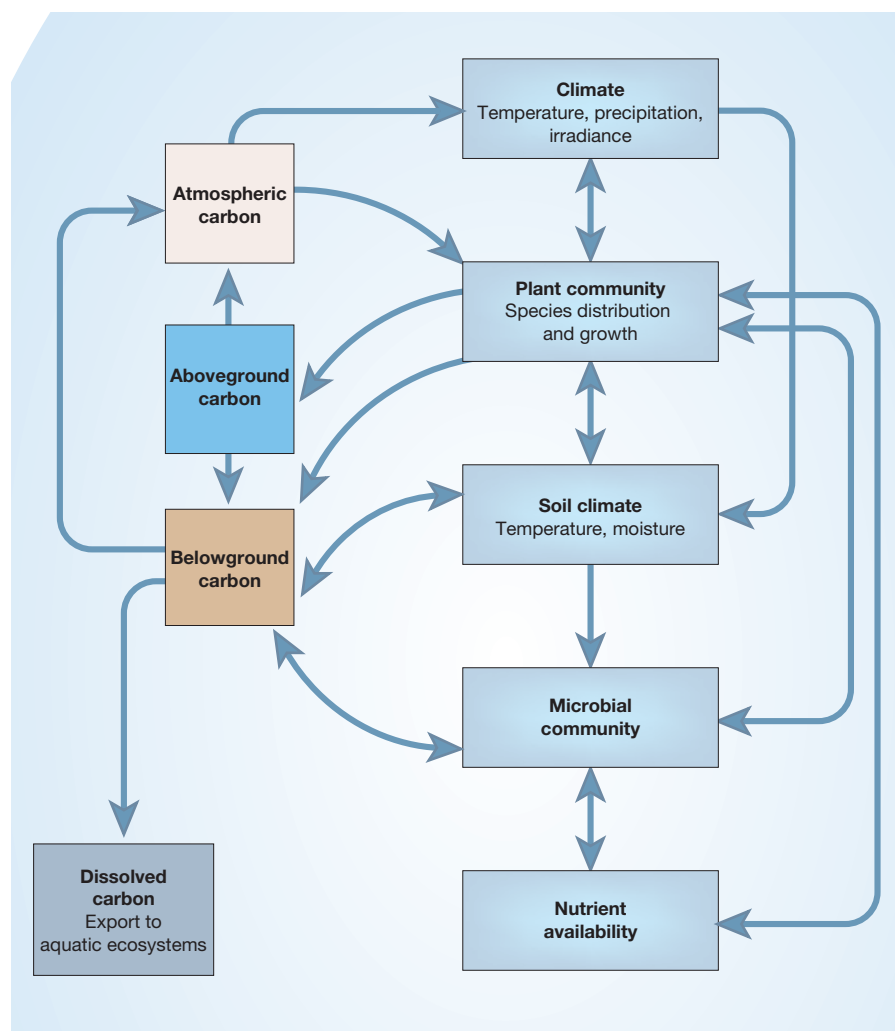


Figure 1 Carbon connections in high-latitude terrestrial ecosystems. On the left are the main pools in which carbon is stored; on the right are the factors that determine the exchange of carbon between those pools. The interactions tested experimentally by Mack *et al.*¹ in their fertilization experiments are those between nutrient availability and the plant community, and the consequences for carbon storage above and below ground.

other hand, warming-induced decomposition could also release nutrients that stimulate plant growth, potentially allowing plants to store more carbon and offset losses of soil carbon.

Predicting the direction of change is complicated by factors that may act independently of one another or may interact to control the carbon balance in high-latitude ecosystems (Fig. 1). Some observations³ suggest that the natural balance of carbon storage and carbon losses is variable — under some conditions the tundra stores carbon; at other times there is a net loss. But it is difficult to predict long-term patterns without decades of monitoring or experimental manipulations that accelerate natural processes.

Mack *et al.*¹ present insights from using both approaches. For more than two decades they have investigated the effects of adding nitrogen and phosphorus to test how nutrient availability controls the carbon balance of moist acidic tundra, a circumpolar

ecosystem type. By constructing a carbon budget for both fertilized and unfertilized plots, Mack *et al.* confirm observations that increased nutrient availability increases carbon storage in plants through accelerated growth of woody shrubs⁴. They also found greater accumulation of carbon in standing dead material and leaf litter on the soil surface, as well as in roots and organic matter in the upper layers of the soil.

However, after digging a little deeper, they discovered that the lower layer of organic matter decreased in thickness, and that there was less carbon in the underlying mineral soil. Also, compared with controls, they found fewer roots growing in the deep soils in the fertilized plots, although total root growth was estimated to be the same. The increase in carbon storage in plants and at the soil surface could not offset the losses deep down, and Mack *et al.*¹ calculate that fertilization caused an overall net carbon loss of 2 kg m^{-2} within two decades, as carbon below ground decreased from about

9 kg m^{-2} in control plots to about 7 kg m^{-2} in fertilized plots.

So where did the carbon go? Although the pathways through which the carbon escaped were not measured directly, the physical reduction in organic matter clearly indicates a loss of carbon from the soil. The authors hypothesize that increased nutrient availability stimulated the activity of decomposer microorganisms, which then proceeded to consume the older, deeper organic matter. If so, these data challenge the dogma that soil microbial activity is generally limited by the availability of carbon⁵, and they support research indicating that microbes in subsurface soil layers may be constrained more by nutrients than by carbon⁶. In addition, some carbon was probably leached away by melting snow or by rainfall, possibly affecting aquatic ecosystems. Microbial activity, in the form of nitrification and denitrification, and nitrate leaching, would also explain where some of the nitrogen has gone — despite inputs of $10 \text{ g m}^{-2} \text{ yr}^{-1}$, the researchers did not find more nitrogen in the fertilized tundra.

The implications of this fertilization experiment, according to Mack *et al.*, are that predicted increases in nutrient availability associated with decomposition of organic matter under warmer temperatures could prime further losses of organic matter. If that loss is accounted for by release of CO_2 into the atmosphere, there would be an additional positive-feedback effect, with enhanced greenhouse-gas-induced warming further increasing decomposition until soils are depleted of carbon and nitrogen. The results question the hypothesis that increased nutrient availability could increase total ecosystem carbon storage through enhanced above-ground productivity⁷. And they are alarming in that the Arctic has been warming at accelerated rates during the past three decades⁸, coincident with an increase in shrub expansion⁹ that possibly signifies that regional carbon losses may already be occurring.

However, it is necessary to remember that this is not a direct observation of warming-induced changes. Among other methodological limitations¹⁰, the direct addition of nutrients in the mineral forms most readily taken up by plants means that microbes are cut out of the initial decomposition process that would lead to the theoretical release of these nutrients. Thus, the natural pathways of carbon and nutrient cycling (Fig. 1) have been bypassed. How an experimentally manipulated supply of nutrients compares with a warming-induced increase is unknown. Moreover, interactions between changing climatic variables — temperature, precipitation, irradiance and higher CO_2 levels — may or may not result in the accelerated decomposition of organic matter or shifts in species composition. In experiments where air temperatures have been warmed by constructing greenhouses on the tundra,

the plant response is not as dramatic as it is with high levels of fertilization¹¹.

Nonetheless, the authors have demonstrated that long-term fertilization resulted in net carbon loss from the tundra ecosystem, probably as a result of increased decomposition of soil organic matter. Research to find out whether decomposition is stimulated directly by increased nutrient availability, or indirectly through changes in the plant or microbial community, will help to identify the pathways involved and to link experiments with forecasts of global climate change.

Finally, these findings add to a growing body of evidence suggesting that increases in aboveground carbon storage through woody-plant encroachment may be offset by losses of soil carbon^{12,13}. Here again, we have a reminder of the need for accurate assessments of carbon storage, both above and below ground, in order to predict ecosystem response to environmental change.

Wendy M. Loya is in the Ecosystem Science Center,

School of Forestry and Environmental Science, Michigan Technological University, Houghton, Michigan 49931, USA.

e-mail: wmloya@mtu.edu

Paul Grogan is in the Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada.

e-mail: groganp@biology.queensu.ca

1. Mack, M. C., Schuur, E. A. G., Bret-Harte, M. S., Shaver, G. R. & Chapin, F. S. III *Nature* **431**, 440–443 (2004).
2. Post, W. M., Emmanuel, W. R., Zinke, P. J. & Strangenburger, A. G. *Nature* **298**, 156–159 (1982).
3. Oechel, W. C. *et al. Nature* **406**, 978–981 (2000).
4. Shaver, G. R. *et al. Ecology* **82**, 3163–3181 (2001).
5. Schimel, J. P. & Weintraub, M. N. *Soil Biol. Biochem.* **35**, 549–563 (2003).
6. Fierer, N., Allen, A. S., Schimel, J. P. & Holden, P. A. *Glob. Change Biol.* **9**, 1322–1332 (2003).
7. Shaver, G. R. *et al. BioScience* **50**, 871–882 (2000).
8. Comiso, J. C. *J. Clim.* **16**, 3498–3510 (2003).
9. Sturm, M., Racine, C. & Tape, K. *Nature* **411**, 546 (2001).
10. Hobbie, S. E., Nadelhoffer, K. J. & Hogberg, P. *Plant Soil* **242**, 163–170 (2002).
11. Van Wijk, M. T. *et al. Glob. Change Biol.* **10**, 105–123 (2003).
12. Jackson, R. B., Banner, J. L., Jobbágy, E. G., Pockman, W. T. & Wall, D. H. *Nature* **418**, 623–626 (2002).
13. Hartshorn, A. S., Southard, R. J. & Bledsoe, C. S. *Soil Sci. Soc. Am. J.* **67**, 1572–1581 (2003).

Evolution

A is for adaptation

Jef D. Boeke

Studies of a bacterial virus have revealed an unexpected weapon that helps it to overcome its host's rapidly changing defences. A look at other organisms hints that the mechanism might be widespread.

Adapt or die. This axiom has been used so many times, in so many different contexts, that its origin is difficult to trace. But it aptly describes an intriguing mechanism exploited by viruses that infect *Bordetella* bacteria. This mechanism guarantees the viruses' survival in the face of host adaptations that would otherwise severely limit their ability to infect and multiply. On page 476 of this issue, Doulatov *et al.*¹ describe how these viruses use a form of the enzyme reverse transcriptase (RT) to generate variability very specifically in the gene encoding a protein required for binding to the bacterial surface. RTs use RNA sequences as templates to generate complementary DNA (cDNA). In this case, it is proposed that the cDNA product directs RT-induced mutations to just the right spot in the viral genome. Natural selection then acts on the mutated progeny viruses to select for those 'winners' that can infect bacteria efficiently. This work sheds new light on the pervasive presence of RT genes in many genomes, suggesting another way in which such genes can benefit the organisms in which they reside.

The *Bordetella* genus includes the causative agents of human whooping cough and canine kennel cough. These bacteria have an elaborate system for evading the immune system of the organism they infect

—the nature of their surface proteins is constantly changing. This presents a formidable challenge to any virus (bacteriophage) attempting to infect *Bordetella* cells, because the first step of viral multiplication is attachment to the bacterial cell surface. Such viruses must evolve exceedingly rapidly to keep pace with a dynamic surface structure as their bacterial host undergoes its own infectious cycle. This is evolution on a very fast track indeed.

The bacteriophage BPP-1 has evolved just such a diversity-generating system, which spawns variant progeny bacteriophages possessing altered surface-binding properties at very high rates (at least 0.1% of the progeny can have significantly altered surface properties). Previous work found that this process is facilitated by interactions between the bacteriophage-encoded RT and two closely related sequences, 134 base pairs long, that lie respectively within and near the gene encoding the surface-binding protein (Fig. 1)². These sequences are called the variable repeat (VR) and the tandem repeat (TR). The TR is an unchanging sequence that provides the raw material for generating variability. On the basis of limited information, it was proposed that the TR is transcribed into an as-yet-undefined RNA, which could in turn be copied by the RT in a highly error-prone

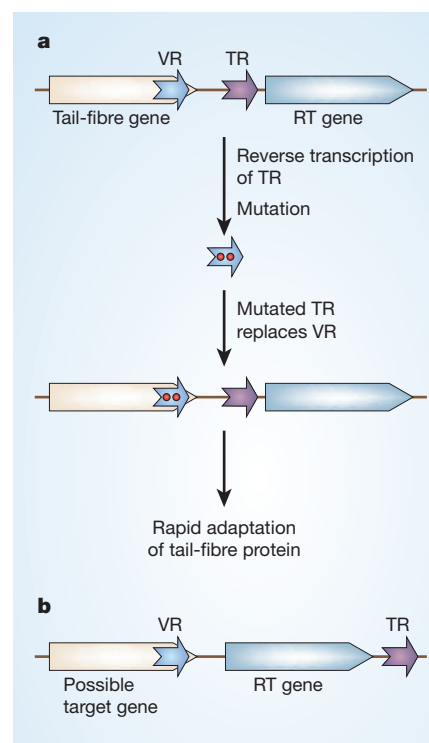


Figure 1 Generating diversity with reverse transcriptase. **a**, Bacteriophage BPP-1, a virus that attacks bacteria of the *Bordetella* genus, contains a reverse transcriptase (RT) enzyme, which, unlike other viral RTs, is not needed for replication. Instead, its role seems to be to create vast diversity within a strictly circumscribed region of the bacteriophage genome, namely the region that encodes the tail-fibre protein, which is needed for the virus to bind to bacteria. This region includes, in addition to the tail-fibre gene, two nearly identical sequences called the variable repeat (VR) and the tandem repeat (TR). TR provides an invariant master source of sequence information. It is copied by RT, during which, as Doulatov *et al.*¹ show, mutations (dots) occur specifically at adenine residues by a process that is not yet understood. The product then replaces a segment of the VR, creating a slightly altered tail-fibre gene. **b**, Doulatov *et al.* also discovered similar VR–TR–RT cassettes in various bacteria, as well as the one shown, from *Nostoc* species (blue-green algae).

manner, generating a hypervariable cDNA product. This would replace a segment of the VR in the genome, producing a swarm of viruses with altered VR sequences in the next generation. So, this VR–TR–RT 'cassette' would maintain TR in a pristine state, while generating diversity in VR.

Doulatov and colleagues¹ now provide more support for this model, and fill in some of the gaps. They show that, remarkably, the diversity-generating system mutates only adenine (A) bases in the TR: by replacing one of the three possible non-A bases in TR with A, they found that the previously invariant base is converted to a hypervariable position in VR. The 23 A residues in the TR sequence

can change to any of the other three bases, generating about 10^{12} different potential versions of VR. The biochemical basis of the specificity for mutation at A is unknown, but presumably the bacteriophage RT could be 'sloppily by design' whenever an A is copied.

Insightful genetic analyses by Doulatov *et al.* also demonstrate that various genetic-selection protocols can produce bacteriophages that have undergone the diversity-generation process. The selection criteria for the altered bacteriophages include natural ones, such as being able to infect a host cell with an altered surface, and artificial ones, such as surviving DNA cleavage by a restriction enzyme. Depending on the nature of the selection, different patterns of survivor mutations were observed.

Restriction enzymes cleave DNA in strictly defined, short sequences. So it is easier to identify and interpret the specific mutations that enable bacteriophages to survive this threat (that is, mutations that prevent DNA cleavage) than to identify those that allow infection of bacteria with altered surface properties. In addition to mutations that occurred directly in the specific restriction-enzyme-recognition site, each variant typically contained a 'patch' of additional mutations flanking the site. However, the patterns of inherited mutation differed dramatically

in the survivors of two different restriction-site selections, with patches centred precisely on the enzyme-recognition sites. Thus, it is the type of selection imposed that apparently determines the pattern of mutations in survivors. These results are consistent with a mechanism in which segments of the variable cDNA are randomly directed to replace the existing VR segment (but not the TR segment).

Reverse transcriptases are central to many RNA viruses, or retroviruses, where they are key to the viral replication process, through an RNA→DNA→RNA mechanism. Being error-prone is an intrinsic property of RTs, which lack the ability to 'proof-read' mistakes made in the copying reaction³ — so the RTs also generate genetic diversity in the viral sequences. Such diversity enables the virus to evolve resistance to host defence mechanisms and drugs. However, there is a crucial and noteworthy difference between the *Bordetella* bacteriophage RT and that of all true retroviruses. The *Bordetella* bacteriophage has a large DNA genome, and it is predicted to replicate by a DNA→DNA mechanism. The bacteriophage RT is therefore completely dispensable for replication — mutants lacking it are merely unable to generate variants. Thus, it appears that the sole 'purpose' of the bacteriophage RT gene

is to create sequence diversity. Perhaps similar logic can explain the long-sought 'function' of another class of RTs, the retrons⁴, long known to generate peculiar but very specific nucleic-acid species through reverse transcription⁵.

The success of this diversity-generating system is underlined by Doulatov and colleagues' discovery that similar cassettes occur in various bacteria and blue-green algae¹. These cassettes all consist of an RT gene and suitably oriented TR and VR sequences, one of which is mutated at the A residues relative to the other. This widespread structural conservation suggests that RT sequences lacking any apparent replicative capability can provide a strong selective advantage, based simply on their ability to generate high levels of DNA diversity within specific genes.

Jef D. Boeke is at the High Throughput Biology Center, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.
e-mail: jboeke@jhmi.edu

1. Doulatov, S. *et al.* *Nature* **431**, 476–481 (2004).
2. Liu, M. *et al.* *Science* **295**, 2091–2094 (2002).
3. Gabriel, A., Willems, M., Mules, E. H. & Boeke, J. D. *Proc. Natl Acad. Sci. USA* **93**, 7767–7771 (1996).
4. Rest, J. S. & Mindell, D. P. *Mol. Biol. Evol.* **20**, 1134–1142 (2003).
5. Lampson, B., Inouye, M. & Inouye, S. *Prog. Nucleic Acid Res. Mol. Biol.* **67**, 65–91 (2001).

Biological techniques

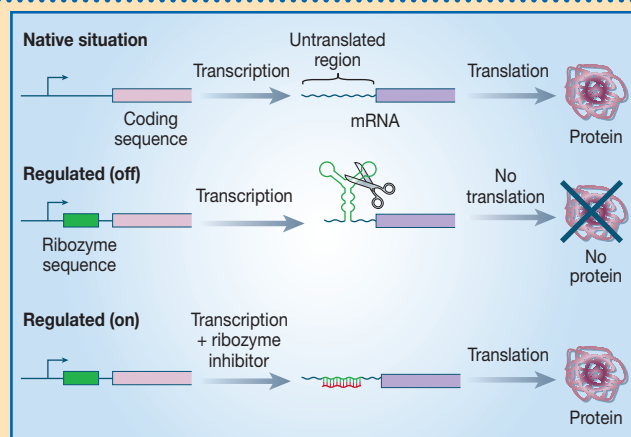
Tailor-made riboswitches

Studies of certain model systems — bacteria, yeast, flies and worms — that can be manipulated genetically have been extremely useful in outlining fundamental molecular signalling pathways that are common to most organisms. But the genomes of more complex creatures, including mammals, can contain an order of magnitude more genes, many of which do not resemble those known from the model systems. Without a means of directly regulating the expression of these genes, it has been difficult to dissect their roles, particularly if they are involved in development- or tissue-specific processes. That looks set to change, however, thanks to an exciting new technique described by Laising Yen and colleagues elsewhere in this issue (*Nature* **431**, 471–476; 2004).

The discovery of ribozymes — RNA sequences that can mediate their own cleavage, in the absence of any protein cofactor — showed that RNAs had the potential to act as regulatory molecules. This potential

was realized with the identification of riboswitches, messenger RNAs containing specialized ribozyme sequences whose self-cleavage is prevented by the binding of a small molecule. Natural riboswitches that recognize ligands such as metabolic intermediates or nucleotides have been described in bacterial systems, where they function to regulate gene expression.

Yen and colleagues exploited this information to develop a way of regulating gene expression in mammalian cells (see illustration). The system consists of two parts: a ribozyme sequence that is both highly efficient and constitutively active *in vivo*, and a means of modulating the ribozyme's self-cleavage activity (using either a ligand or an oligonucleotide complementary to the ribozyme sequence). If a ribozyme-encoding DNA sequence is placed upstream of the coding region of a gene of interest, the gene's expression will be repressed unless the small inducer molecule or antisense



oligonucleotide is provided.

Needless to say, a certain amount of fiddling was required to find a workable ribozyme-inducer pair. Once done, however, the system could regulate gene expression in many cultured cell types and in mice; expression was almost undetectable unless the inducer molecule was present. Proof of this principle was obtained with a transgene that the authors inserted into the cells.

In conjunction with other emerging technologies, such as

aptamers, it may be possible to tailor gene-regulation systems to respond to virtually any small molecule or metabolite. If ribozyme sequences are incorporated into an endogenous gene, it should also be possible to monitor that gene's function during development or in specific tissues by supplying specific concentrations of intracellular molecules. And having different ribozyme sequences will allow several genes to be examined either singly or together.

Angela K. Eggleston



100 YEARS AGO

Scientific critics in Berlin are now much exercised with regard to the remarkable performances of "Clever Hans," the thinking horse. According to the daily Press, a representative committee... witnessed these performances with the view of ascertaining whether they were the result of a trick, or whether they were due to the mental powers of the animal. Their verdict, it is reported, was unanimous in favour of the latter view. It is stated that when told that the day was Tuesday, and asked which day of the week this represented, the horse would give the correct answer by taps. Similarly he will tell not only the hour, but the minutes indicated by a watch.

From *Nature* 22 September 1904.

50 YEARS AGO

The decision of the Atomic Energy Commission... that Dr. J. R. Oppenheimer should be denied access to restricted data because, on the record before the Commission, he was not entitled to the continued confidence of the Government and of the Commission "because of the proof of fundamental defects in his 'character'" is, of course, a matter of domestic policy within the United States. Even had, however, Great Britain and the United States not been somewhat "mixed up" (Sir Winston Churchill's phrase) in the early development of the atom bomb, the termination in this manner of the Government career of one who had rendered such outstanding services in this field could not but be a matter of profound interest, if not concern, to scientific circles in Britain... Apart from the question whether even the most richly endowed nation can afford to divert an unlimited amount of its man-power to security investigations, the contrast between the ambiguous answers produced after so much voluminous inquiry and the swiftness and perspicacity with which Sir John Woods's committee investigating the Crichton Down episode in Britain reported on issues touching the public interest no less vitally is startling. Almost inevitably it suggests that there may be substance in what some American scientists are saying, and that Dr. Oppenheimer has been dismissed because his opinions were unpalatable to the authorities... who have chosen this retrograde manner of removing an adviser so as to minimize the possibility that others may avail themselves of his advice or services.

From *Nature* 25 September 1954.

Materials science

The art of restoration

David Erhardt

There are various techniques for the restoration of artwork — how effective and safe these are also varies. 'Reversible' gels could, however, provide a less risky way to reverse the ravages of time.

The cleaning of paintings is one of the most controversial activities that can be conducted in a museum. Unlike many artefacts, much of the value of a painting (both aesthetic and monetary) depends on its appearance, which in turn depends on the condition of its surface, especially the first few micrometres. The removal of dirt, yellowed varnishes and later overpaint is a critical step in the restoration of a painting, to return it to as close to its original state and appearance as possible. In *Langmuir*, Carretti *et al.*¹ report the development of 'reversible' gels for the removal of non-original layers. These solvent-containing gels can be converted back to a free-flowing fluid state *in situ* for easier and more complete removal of the cleaning mixture. This work illustrates both the value of applying science to the preservation of artwork and the necessity of understanding the practical problems involved in determining the safety of any restoration technique.

The problem lies not in removing unwanted material from the surface of a painting, but in doing so without removing, disturbing or otherwise altering the original design layer. This can be extremely difficult, especially if the solubility of the overlying layer is similar to that of the paint binder (medium) — or, even worse, if the layer to be removed is less soluble than the original layer in any reagent. Traditionally, reagents such as solvents, spirits, alkalis, acids and soaps have been used, as well as simple mechanical action (scraping with a scalpel). More recently, scientific approaches have been brought to bear: solvent theory, explaining the action of solvents, has taken much of the guesswork out of preparing solvent mixtures of the desired strength; enzymes can remove very specific types of materials (proteases for proteins, lipases for oils, and so on). Lasers can simply vaporize the layer to be removed if it is otherwise intractable or if the original layer is too sensitive to other reagents; concurrent spectroscopic monitoring of the plasma plume of material removed by the laser indicates when the interface between different layers has been reached². Still, the cleaning of paintings relies as much on the skill, experience and judgement of the conservator as it does on science, as any technique improperly applied can do damage.

In recent years, a number of gelled reagents have been introduced³. These polymer-based materials make it easier to



Figure 1 Brighter future? This fourteenth-century icon, held at the National Gallery of Siena, Italy, has become dulled over time. Tests on a small area (not shown) by Carretti *et al.*¹ suggest that 'reversible' gels could be effective cleaning agents that would not damage the artwork.

localize the action of cleaning, to combine the action of otherwise immiscible components, and to avoid the problem of quick evaporation of volatile solvents. Unfortunately, much of the literature on these materials consists of anecdotal case studies that have not been peer-reviewed and which often contain unsubstantiated, even incorrect, statements regarding their safety, specificity, mechanism of action and the roles of the individual components. The safety of gelled reagents has not yet been adequately demonstrated, and even more serious problems than those with volatile solvents have been documented⁴.

One major problem is the difficulty of completely removing the viscous gelled reagent and any non-volatile materials such as soaps or high-boiling-point liquids contained in the mixture. This is especially true when the material has been forced into cracks or a porous layer during the cleaning process. But Carretti *et al.*¹ have an ingenious approach: their gelled solvent mixture reverts to a free-flowing liquid on addition of a small amount (a few microlitres) of a weak

acid solution (0.05 M acetic acid). The resulting 'degelled' mixture is more easily and completely removed than are viscous gels, in this case by absorption into a cotton swab. Further clearance of any remaining residue is facilitated by the fact that the gelling agent is now readily soluble.

Working on a small area of a fourteenth-century icon (Fig. 1) from the National Gallery in Siena, Italy, Carretti *et al.*¹ have demonstrated that their rheoreversible gels can remove material from the surface of a painting. The spectrum of X-rays absorbed by the removed material indicated that no mercury was present, which implies that none of the vermilion pigment (mercuric sulphide) of the original paint layer had been removed. However, removal of inorganic pigments is not often the major problem, because most are insoluble in the standard cleaning agents. More serious is the extraction of organic material such as organic dyes and pigments, or the low-molecular-weight compounds in the medium that function as plasticizers in maintaining the flexibility of the paint film.

Carretti and colleagues' infrared analysis

of the removed material showed that the varnish was a natural resin. But infrared analysis does not find the minor amounts (relative to varnish) of paint media that are typically extracted. Infrared spectra of the common resin varnishes are quite variable, and also change with age, making the interpretation of resin spectra difficult to begin with, and the detection of minor components even less likely. Methods such as gas chromatography are much more useful in this context for evaluating the amount of paint media removed (with chromatography it is quite easy to look for compounds, such as fatty acids in oil paint, that are specific to the paint rather than to the resin). Similarly, infrared analysis of the cleaned surface cannot adequately demonstrate that no traces of cleaning mixture remain on the surface. Even scanning electron micrographs of the surface that show a surface similar to an uncoated paint film cannot unequivocally demonstrate that components of the paint film have not leached out.

The development and testing of new cleaning techniques and reagents for the restoration of paintings is a daunting task.

Both techniques and reagents must be shown to be safe, causing no significant alteration of, or damage to, the original layers of the artwork. And this must be proved before advocating or promoting their use, which has not always been done by other researchers. Carretti *et al.*¹ have made significant progress in developing their new method, but much still remains to be done in evaluating the results and refining the technique to minimize any undesired effects. The rheoreversible gels developed by this team are a positive step in the development of gelled cleaning mixtures that are safer than those presently in use. The field should look forward to their further contributions on the subject. ■

David Erhardt is at the Smithsonian Center for Materials Research and Education, Smithsonian Institution, Museum Support Center, 4210 Silver Hill Road, Suitland, Maryland 20746-2863, USA. e-mail: erhardtd@scmre.si.edu

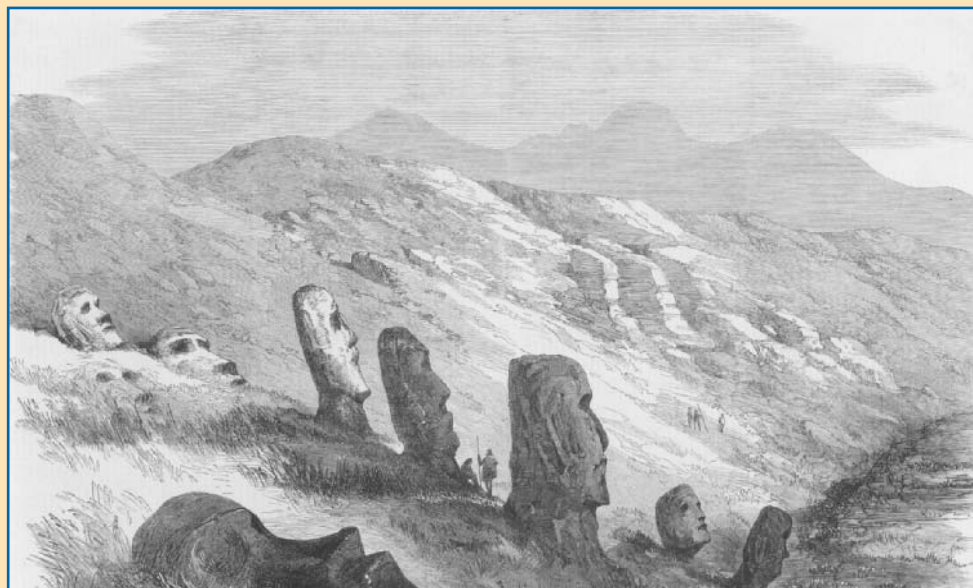
1. Carretti, E., Dei, L., Macherelli, A. & Weiss, R. G. *Langmuir* doi:10.1021/la0495175 (2004).
2. Scholten, J. H., Teule, J. M., Zafropoulos, V. & Heeren, R. M. A. *J. Cultural Heritage* 1, 215–220 (2000).
3. Wolbers, R. *Cleaning Painted Surfaces, Aqueous Methods* (Archetype, London, 2000).
4. Erhardt, D. & Bischoff, J. *Studies Conserv.* 39, 3–27 (1994).

Environmental geography

Treeless at Easter

Easter Island in the eastern Pacific is one of the remotest spots on Earth, but distance need not lend enchantment to the view. When Easter was discovered by the Dutch explorer Jacob Roggeveen in 1722, he found it a treeless wasteland rather than the palm-fringed paradise one usually associates with the Pacific. Captain James Cook (writing in 1774) described the islanders as "small, lean, timid and miserable", hanging on as subsistence farmers amid the ruins of the giant statues erected by their ancestors (see picture).

The statues were erected between the eleventh and seventeenth centuries, and could have been the immediate cause of the islanders' plight. At its height, Easter Island society was based on a system of clans, who outdid one another in feats of megalithic excess. The strain cost the island all its native birds and all but a few of its native trees, which included the tallest species of palm tree in the world. Having turned their island's natural capital into artefacts, the islanders relapsed into war, savagery and cannibalism. In "Twilight at Easter", an article in



The New York Review of Books (25 March 2004), Jared Diamond tells the story of Easter Island as a tragic parable for modern times.

But the islanders were, in addition, cursed by poor location. Elsewhere in this issue (*Nature* 431, 443–446; 2004), Barry Rolett and Diamond present an analysis of environmental factors that might be associated with the deforestation of Pacific islands. They show that Easter had

drawn a losing hand even before the first Polynesian colonists stepped ashore.

Islands most likely to lose their forests are small, dry, remote from other islands (and from continental dust inputs), low-lying and relatively distant from the Equator. Easter scores high on all these factors. "Easter's collapse was not because its people were especially improvident but because they faced one of the Pacific's most fragile environments," according to Rolett

and Diamond. Or, in the words of the blues standard: "If it wasn't for bad luck, I wouldn't have no luck at all." In the final analysis, megalithomania was probably the last straw. Easter Island's current environmental profile cannot be wholly explained by natural factors, as Rolett and Diamond's model shows — and neither can the state of relatively well-wooded Pacific Islands such as Tonga, whose society employs its own protective measures against deforestation. **Henry Gee**

Neurobiology

Nerves of rubber stretched out

J. Neurosci. **24**, 7978–7983 (2004)

The long projections of nerve cells, known as axons, must lengthen to keep pace with a growing animal — and in the spine of the blue whale this happens at a rate of up to 3 cm per day. But until now, researchers lacked much experimental evidence for how axons achieve such a feat.

Bryan J. Pfister *et al.* custom-built an axon-stretching device in which two populations of neurons were seeded on adjacent membranes. Connecting these neurons were axons, about 0.1 mm in length, that were gradually stretched by cranking the membranes apart with a computer-driven motor. Under the right conditions, the stretched axons could be grown at rates of up to 8 mm per day, and reach lengths of up to 10 cm without snapping or compromising the workings of their organelles. The image reproduced here shows two sets of neurons (top and bottom) connected by 5-cm-long bundles of stretch-grown axons.

The researchers have yet to work out how the cell transports and inserts building materials to achieve this feat of stretching. But they propose that the system they have devised might be used to grow long axonal tracts and bridge areas of damage for nerve repair.

Helen Pearson



^7Be has a half-life of roughly 53 days 3 hours, the enclosed atoms have a half-life that is about 10.5 hours shorter.

Mark Peplow

Cell biology

A protein digestif

Dev. Cell **7**, 359–371 (2004)

The pancreas is well known for its ability to regulate blood sugar levels, but the bulk of its mass functions to secrete enzymes into the gastrointestinal tract to help us digest our food. The factors governing this release are uncertain, but Cheng-Chun Wang *et al.* find that a protein called VAMP8 (or endobrevin) may be involved.

They discovered VAMP8 in the membrane of zymogen granules, tiny enzyme-filled sacs that are found in most pancreatic cells. They also found that pancreatic cells from mice lacking VAMP8 contain three times more zymogen granules than controls — yet the granules are unable to spill their contents. This suggests that VAMP8 helps to regulate enzyme release from these specialized pancreatic cells.

The authors think that the protein might also be involved in pancreatitis (inflammation of the pancreas), which occurs when digestive enzymes are activated in the pancreas and are secreted abnormally into the blood. Mice lacking VAMP8 are partially resistant to the condition and have lower blood levels of digestive enzymes than do controls with pancreatitis.

Helen Pilcher

Chemical biology

A molecular stress gauge

Angew. Chem. Int. Edn **43**, 4750–4752 (2004)

Proteins that bind to DNA often bend the double helix. The energy needed for this mechanical distortion comes from the favourable binding interaction between protein and DNA. Wanqiu Shen *et al.* have constructed a DNA molecular device that measures how much excess binding energy a protein has available for distorting DNA.

The device contains a non-covalent tether that prevents bending; the authors vary the strength of the tether until the protein has enough excess binding energy to snap it. They use a protein from the bacterium *Escherichia coli* called IHF, which bends DNA when it sticks to a recognition sequence on the DNA molecule. In the measuring device this sequence is sandwiched between two rigid units labelled with fluorescent dyes. When IHF binds, bending causes the dyes to swivel apart, decreasing the resonant energy transfer between them — this is detected spectroscopically. The two end units are hooked together by complementary single-stranded DNA with a variable number of overlapping base pairs: the

shorter the overlap, the weaker the tether.

The researchers find that bending always happens for a strand disruption energy of less than 3 kcal per molecule, and is always prevented when this energy exceeds 7–8 kcal per molecule. In between, there seems to be a mixture of disrupted and intact DNA devices: binding and bending by IHF is marginally possible.

Phillip Ball

Nuclear physics

Accelerated decay

Phys. Rev. Lett. **93**, 112501 (2004)

The radioactive decay of beryllium-7 (^7Be) speeds up when individual atoms are trapped inside the cage of a fullerene molecule (C_{60}), according to T. Ohtsuki and colleagues. An atom's environment is known to affect its half-life, but the influence of C_{60} on ^7Be seems to be the greatest ever recorded — the half-life of the decay changes by about 0.8%.

The ^7Be atom can decay by the capture of one of its own electrons by its nucleus, a process in which a proton is transformed into a neutron. But when the ^7Be atom is inside a C_{60} molecule, there is considerable overlap of the electron density of the atom and that of its carbon cage, which makes electron capture more likely.

The authors report that, whereas metallic

Physics

Water-wave regulation

Appl. Phys. Lett. **85**, 1645–1647 (2004)

Crystals that are composites of different materials can be engineered to block certain wavelengths of light: as light passes from one material to another, scattering of the waves effectively weeds out specific wavelengths. Likewise, theory predicts that an array of solid structures should have the same effect on water waves.

Taek Seong Jeong *et al.* have provided a practical demonstration of this phenomenon, in an experiment in which water waves passed through an array of submerged acrylic cylinders fixed to the bottom of a shallow water tank. The cylinders were laid out in different patterns — a hexagonal arrangement (which looks rather like the atomic structure of graphite) cut out water wavelengths from about 47 mm to 68 mm, while a triangular pattern damped down waves in the 23–38-mm range.

The authors point out that an offshore array of cylinders could diffuse waves that contribute to coastal erosion. They calculate that an arrangement of cylinders in a series of 10-metre-wide hexagons could effectively block coastal water waves that have a period of a few seconds.

Mark Peplow

Morphine-pathway block in *top1* poppies

This opium poppy mutant provides chemical precursors for non-addictive analgesics.

The opium poppy is a source of the pharmaceuticals codeine, morphine and their derived analgesics. Here we describe the initial characterization of the poppy mutant known as *top1* (for ‘thebaine oripavine poppy 1’), which accumulates the morphine and codeine precursors thebaine and oripavine and does not complete their biosynthesis into morphine and codeine. The original discovery of *top1* stimulated a re-engineering of the opioid industry in the island state of Tasmania, which grows over 40% of the world’s licit opiates, in order to produce thebaine and oripavine efficiently from morphine-free poppy crops to provide precursors for highly effective analgesics and for treatment of opioid addiction.

The morphinan branch of the normal poppy pathway, from salutaridine to morphine, involves two demethylation steps and proceeds by two alternative routes (Fig. 1)^{1,2}. Although the enzymatic synthesis of morphine has been almost completely elucidated³⁻⁵, knowledge about the genes involved in the morphinan branch, including their coordination and regulation, is sparse.

We treated seeds of a commercial poppy cultivar (*Papaver somniferum*) with a mutagen and then screened the progeny plants.

The mutant *top1* was found to accumulate thebaine and oripavine but not morphine or codeine (Fig. 2; for details of methods, see supplementary information). The segregation of 375 F₂ individuals was mendelian and was consistent with the involvement of a single gene.

The only visible phenotypic change in the mutant is that its latex is pigmented rather than the normal white. The latex colour change co-segregates with the thebaine phenotype as a reliable visual marker. Feeding of radioactive intermediates (see supplementary information) confirmed that there is a block in both arms of the bifurcated pathway at thebaine and oripavine in *top1*. This block may be due to a defect in the enzyme thebaine demethylase, which is likely to be responsible for the 6-*O*-demethylation of both thebaine and oripavine.

We developed a 17,000-gene poppy microarray and used it to explore global changes in gene expression in the mutant. Ten genes were identified as being significantly differentially underexpressed in *top1*, a result verified by quantitative real-time polymerase chain reaction; seven of the clones were highly homologous to known genes of other species (for details, see supplementary information). These include a component of the signal-recognition particle that mediates protein trafficking, a flippase ATP-dependent transmembrane transporter, and a homologue to ftsH protein, a transmembrane ATP-dependent metalloprotease. Three other clones encoded proteins with similarity to known enzymes: phosphoenolpyruvate carboxykinase, aspartate aminotransferase and aldose 1-epimerase.

A comparison of latex proteins revealed many obvious changes in *top1* in both the serum and membrane fractions (results not shown). It seems that the *top1* mutation triggers a series of changes in gene transcripts, proteins and secondary products. None of the candidate genes from the microarray study has yet been proved to be

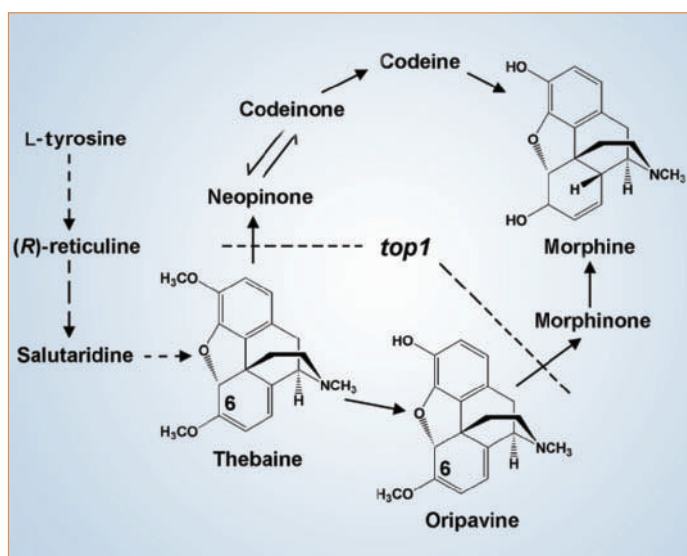


Figure 1 An abbreviated biosynthesis pathway, leading from tyrosine to morphine and showing the two alternative routes from thebaine to morphine. (*S*)-reticuline is a major branchpoint precursor of a number of other alkaloids found in *Papaver somniferum* and in other plant species. The blocks in the pathway in the mutant *top1* are indicated by dotted lines through the enzymatic steps.

a demethylase or otherwise responsible for the *top1* phenotype. Although the apparently coordinated changes in these transcripts and proteins are not understood, a window is opening into the complex metabolic network associated with the morphine pathway and new avenues for modifying the pathway are emerging.

There are several possible explanations for the *top1* mutation phenotype: the gene encoding the 6-*O*-demethylase that acts on thebaine and oripavine might be affected, as could a gene that regulates its function or expression (such as a transcriptional regulator or a protein in a multiprotein complex). Alternatively, there may be an alteration in a structural or transport component that prevents the substrates thebaine and oripavine from arriving at the subcellular compartment where *O*-demethylation occurs⁶⁻⁸.

The *top1* mutant is not only contributing to the genetic dissection of the morphine pathway, it has been swiftly deployed for an agriculturally viable supply of thebaine and is a crop that carries little risk of diversion for illicit purposes. Its manufactured derivatives include oxycodone, buprenorphine, naloxone and naltrexone. Buprenorphine is a mixed agonist–antagonist and a potent analgesic with a favourable side-effect profile. It is playing an expanding role in both severe pain control and in the treatment of addiction.

Anthony G. Millgate*†, Barry J. Pogson†,
Iain W. Wilson*, Toni M. Kutchan‡,
Meinhart H. Zenk§, Wayne L. Gerlach||,
Anthony J. Fist¶, Philip J. Larkin*

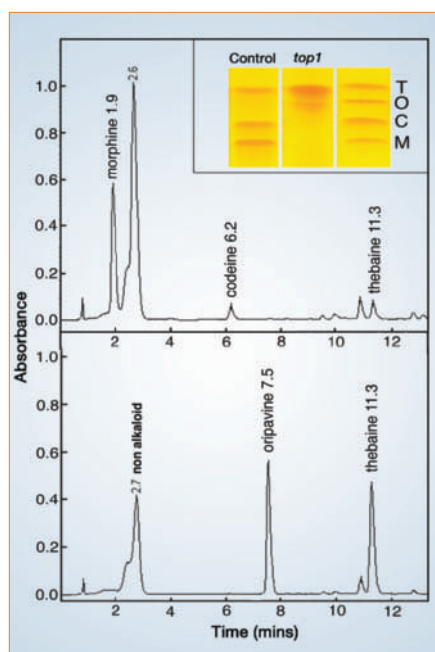


Figure 2 High-pressure liquid chromatography traces of alkaloid profiles from the control parent poppy (upper) and the mutant *top1* (lower). The *top1* trace lacks codeine and morphine peaks. Inset, thin-layer chromatogram of latex from the control and the mutant *top1* (standards in third lane: T, thebaine; O, oripavine; C, codeine; M, morphine). Thebaine and oripavine accumulate as the dominant alkaloids in the mutant at the expense of morphine and codeine, which are evident in the parental phenotype.

*CSIRO Plant Industry, Black Mountain, Canberra, ACT 2601, Australia

e-mail: philip.larkin@csiro.au

†School of Biochemistry and Molecular Biology, The Australian National University, Canberra, ACT 0200, Australia

‡Leibniz Institute of Plant Biochemistry, Department of Natural Product Biotechnology, 06120 Halle (Saale), Germany

§Biozentrum Universität Halle, 06120 Halle (Saale), Germany

||Johnson & Johnson, Locked Bag 4555, Strawberry Hills, New South Wales 2012, Australia

¶Tasmanian Alkaloids, Westbury, Tasmania, 7303 Australia

1. Parker, H. I., Blaschke, G. & Rapoport, H. *Am. Chem. Soc.* **94**, 1276–1282 (1972).
2. Brochmann-Hanssen, E. *Planta Medica* **50**, 343–345 (1984).
3. Kutchan, T. M. *Gene* **179**, 73–81 (1996).
4. Zenk, M. H. in *Organic Reactivity: Physical and Biological Aspects* (eds Maskill, H., Golding, B. T. & Griffin, R. J.) 89–109 (Royal Society of Chemistry, Cambridge, UK, 1995).
5. Kutchan, T. M., Frick, S. & Weid, M. in *Advances in Plant Biochemistry and Molecular Biology* (eds Bohnert, H. J. & Nguyen, H. T.) (Elsevier Science, Oxford, in the press).
6. Bird, D. A., Franceschi, V. R. & Facchini, P. J. *Plant Cell* **15**, 2626–2635 (2003).
7. De-Luca, V. & St Pierre, B. *Trends Plant Sci.* **5**, 168–173 (2000).
8. Bock, A., Wanner, G. & Zenk, M. H. *Planta* **216**, 57–63 (2002).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Ecology

Widespread colonization by polar hypoliths

High-latitude polar deserts are among the most extreme environments on Earth. Here we describe a large and previously unappreciated habitat for photosynthetic life under opaque rocks in the Arctic and Antarctic polar deserts. This habitat is created by the periglacial movement of the rocks, which allows some light to reach their underside. The productivity of this ecosystem is at least as great as that of above-ground biomass and potentially doubles previous productivity estimates for the polar desert ecozone.

The underside of rocks in climatically extreme deserts acts as a refugium for photosynthetic microorganisms (defined as 'hypoliths') and their community (the 'hypolithon')¹. Here, the organisms are protected from harsh ultraviolet radiation² and wind scouring, and trapped moisture can provide them with a source of liquid water³. Colonization also usually requires the rocks to be sufficiently translucent to allow for the penetration of light—all hypoliths reported so far have been found under quartz, which is one of the most common translucent rocks^{4,5}.

We examined 850 randomly selected opaque dolomitic rocks, without regard to the local patterns of periglacial rock sorting, on Cornwallis Island and Devon Island in the

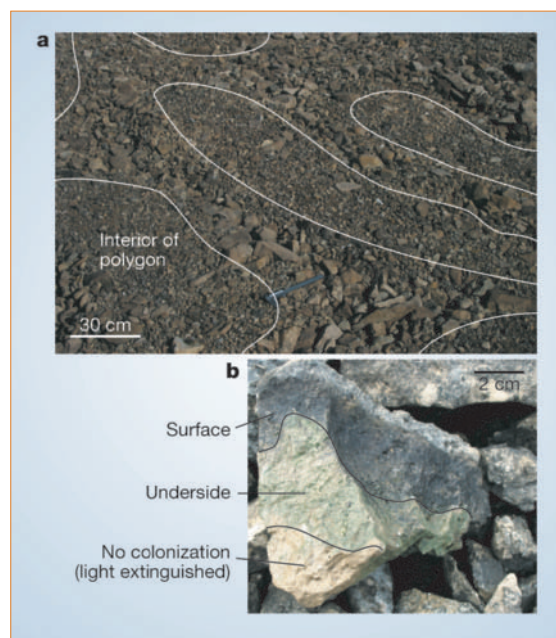


Figure 1 Colonization by polar hypoliths. **a**, Hypolithic colonization is enhanced by rock sorting which, among other factors, increases light penetration to the underside of rocks, shown here at the edges of polygonal terrain (white lines indicate polygon edges). **b**, Opaque rocks are colonized on their underside by well defined bands of photosynthetic communities, seen here on an excavated rock.

Canadian high Arctic. These are regions typical of extreme polar desert, where vegetation cover was measured to be 1.2% or less⁶. On Devon Island, 95% of rocks were found to be colonized, and on Cornwallis Island, 94% were colonized. The photosynthetic organisms form a well defined green band on the underside of the rocks. The mean ($\pm$ s.d.) band width of these communities was 3.1 ± 1.9 cm and 3.0 ± 1.6 cm on Devon Island and Cornwallis, respectively.

The Arctic hypoliths are dominated by cyanobacteria. Species found include *Gloeocapsa* cf. *atrata* Kützinger, *Gloeocapsa* cf. *punctata* Nägeli, *Gloeocapsa* cf. *kuetziniana* Nägeli and *Chroococcidiopsis*-like cells; unicellular algal chlorophytes were also present⁷.

We investigated the colonization of well developed polygons, which are just one manifestation of a diversity of linear and circular features caused by the long-term action of periglacial processes⁸. In the Arctic, we found colonization on 68% of rocks within polygons, with a mean ($\pm$ s.d.) band width of 0.7 ± 0.8 cm, where the rocks are surrounded by fine soil sorted into their centres. At the edges, where the cracks around the rocks are larger, we found 100% colonization with a mean ($\pm$ s.d.) band width of 3.6 ± 1.4 cm (Fig. 1).

We studied similar polygonal terrains at Mars Oasis on Alexander Island in the Antarctic Peninsula, a location where hypolithic colonization occurs. The percentage colonization was 5% within polygons and 100% on the edges of polygons, with mean ($\pm$ s.d.) band widths of 0.7 ± 0.1 and

2.1 ± 0.3 cm, respectively. We propose that rock sorting by periglacial action, including that during freeze–thaw cycles, improves light penetration around the edges of rocks, one factor that might account for the widespread colonization of the underside of opaque rocks.

We estimated the productivity of the Arctic communities by monitoring the uptake of ¹⁴C-labelled sodium bicarbonate (for methods, see supplementary information). Assuming no carbon uptake from other sources, a conservative estimate of mean ($\pm$ s.d.) productivity is about 0.8 ± 0.3 g m⁻² yr⁻¹. Given that the estimated mean ($\pm$ s.d.) productivity from plants, lichens and bryophytes on Devon Island is 1.0 ± 0.4 g m⁻² yr⁻¹ (ref. 6), the polar hypolithon may be just as important in sequestering carbon, at least doubling previous estimates of the productivity of the rocky polar desert.

Charles S. Cockell*, M. Dale Stokes†

*British Antarctic Survey, Natural Environment Research Council, High Cross, Cambridge CB3 0ET, UK

e-mail: cscoc@bas.ac.uk

†Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92037-0238, USA

1. Golubic, S., Friedmann, I. & Schneider, J. *J. Sediment. Petrol.* **51**, 0475–0478 (1981).
2. Cockell, C. S., Rettberg, P., Horneke, G., Scherer, K. & Stokes, D. M. *Polar Biol.* **26**, 62–69 (2003).
3. Smith, M. C., Bowman, J. P., Scott, E. J. & Line, M. A. *Antarctic Sci.* **12**, 177–184 (2000).
4. Broady, P. A. *Brit. Phycol. J.* **16**, 231–240 (1981).
5. Schlesinger, W. H. *et al. Ecology* **84**, 3222–3231 (2003).
6. Bliss, L. C., Svoboda, J. & Bliss, D. I. *Holarctic Ecol.* **7**, 305–324 (1984).
7. Cockell, C. S. *et al. Meteor. Planet. Sci.* **37**, 1287–1298 (2002).
8. Kessler, M. A. & Werner, B. T. *Science* **299**, 380–383 (2003).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

brieF communications arising online

www.nature.com/bca

Corrigendum: Does gut hormone PYY_{3–36} decrease food intake in rodents?

M. Tschöp, T. R. Castañeda, H. G. Joost, C. Thöne-Reineke, S. Ortmann, S. Klaus, M. M. Hagan, P. C. Chandler, K. D. Oswald, S. C. Benoit, R. J. Seeley, K. P. Kinzig, T. H. Moran, A. G. Beck-Sickinger, N. Koglin, R. J. Rodgers, J. E. Blundell, Y. Ishii, A. H. Beattie, P. Holch, D. B. Allison, K. Raun, K. Madsen, B. S. Wulff, C. E. Stidsen, M. Birringer, O. J. Kreuzer, M. Schindler, K. Arndt, K. Rudolf, M. Mark, X. Y. Deng, D. C. Withcomb, H. Halem, J. Taylor, J. Dong, R. Datta, M. Culler, S. Craney, D. Flora, D. Smiley, M. L. Heiman *Nature* **430**, 165 (2004); doi:10.1038/nature02665 (Corrigendum doi:10.1038/nature03019)

Corrigendum**Physiology: Does gut hormone PYY₃₋₃₆ decrease food intake in rodents?**

M. Tschöp, T. R. Castañeda, H. G. Joost,
C. Thöne-Reineke, S. Ortmann, S. Klaus, M. M. Hagan,
P. C. Chandler, K. D. Oswald, S. C. Benoit, R. J. Seeley,
K. P. Kinzig, T. H. Moran, A. G. Beck-Sickinger, N. Koglin,
R. J. Rodgers, J. E. Blundell, Y. Ishii, A. H. Beattie,
P. Holch, D. B. Allison, K. Raun, K. Madsen, B. S. Wulff,
C. E. Stidsen, M. Birringer, O. J. Kreuzer, M. Schindler,
K. Arndt, K. Rudolf, M. Mark, X. Y. Deng, D. C. Withcomb,
H. Halem, J. Taylor, J. Dong, R. Datta, M. Culler,
S. Craney, D. Flora, D. Smiley, M. L. Heiman
(doi:10.1038/nature02665)

The name of one author (D. C. Whitcomb) was misspelled in this Brief Communications Arising.

Corrigendum doi:10.1038/nature03019

Impaired PtdIns(4,5)P₂ synthesis in nerve terminals produces defects in synaptic vesicle trafficking

Gilbert Di Paolo^{1,2}, Howard S. Moskowitz⁵, Keith Gipson³, Markus R. Wenk^{1,2}, Sergey Voronov^{1,2}, Masanori Obayashi^{1,2}, Richard Flavell^{1,4}, Reiko M. Fitzsimonds³, Timothy A. Ryan⁵ & Pietro De Camilli^{1,2}

¹Howard Hughes Medical Institute, Departments of ²Cell Biology, ³Physiology and ⁴Section of Immunobiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA

⁵Department of Biochemistry, Weill Medical College of Cornell University, New York, New York 10021, USA

Phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)P₂) has an important function in cell regulation both as a precursor of second messenger molecules and by means of its direct interactions with cytosolic and membrane proteins. Biochemical studies have suggested a role for PtdIns(4,5)P₂ in clathrin coat dynamics, and defects in its dephosphorylation at the synapse produce an accumulation of coated endocytic intermediates. However, the involvement of PtdIns(4,5)P₂ in synaptic vesicle exocytosis remains unclear. Here, we show that decreased levels of PtdIns(4,5)P₂ in the brain and an impairment of its depolarization-dependent synthesis in nerve terminals lead to early postnatal lethality and synaptic defects in mice. These include decreased frequency of miniature currents, enhanced synaptic depression, a smaller readily releasable pool of vesicles, delayed endocytosis and slower recycling kinetics. Our results demonstrate a critical role for PtdIns(4,5)P₂ synthesis in the regulation of multiple steps of the synaptic vesicle cycle.

Inositol phospholipids (phosphoinositides) have pleiotropic functions in cell regulation^{1–4}. It has long been known that PtdIns(4,5)P₂, a phosphoinositide concentrated in the plasma membrane, regulates synaptic transmission through the modulatory effects on Ca²⁺ signalling of its metabolic products inositol trisphosphate and diacylglycerol (DAG)^{5,6}. In addition, recent studies have suggested that PtdIns(4,5)P₂ itself and its lipid metabolite DAG may have a direct function in the exo- and endocytosis of synaptic vesicles^{7–13}.

Synaptic vesicles dock at the active zone of synapses, where they undergo a priming step that makes them competent for calcium-dependent fusion. After exocytosis, their membranes are rapidly retrieved from the cell surface in a reaction that requires the GTPase dynamin, and they are re-used to generate new synaptic vesicles¹⁴. A major pathway through which this recycling process occurs is clathrin-mediated budding^{15,16}, although a faster parallel recycling pathway ('kiss-and-run') has also been proposed^{17–19}. After fission, clathrin-coated vesicles rapidly shed their coat and are directed back to the vesicle cluster, from which they can be mobilized for subsequent rounds of release¹⁴.

A direct role for PtdIns(4,5)P₂ in synaptic vesicle endocytosis was suggested by two convergent sets of results. First, a PtdIns(4,5)P₂ phosphatase, synaptojanin 1, was identified as a protein neighbour of dynamin²⁰ and shown to be required for a normal synaptic vesicle cycle^{8,10,11,21,22}. Second, endocytic clathrin adaptors and other endocytic factors were reported to bind lipids²³ and to contain PtdIns(4,5)P binding sites^{3,24,25}. PtdIns(4,5)P₂ cleavage by synaptojanin 1 may help to shed endocytic proteins after endocytosis^{8,10,11,21,22}. PtdIns(4,5)P₂ was also implicated in exocytosis by studies of peptide-containing granules from neuroendocrine cells^{3,26,27}, although it remains unclear whether PtdIns(4,5)P₂ also has a direct physiological role in the exocytosis of synaptic vesicles^{7,13,28}.

The major pathway for the generation of PtdIns(4,5)P₂ involves the phosphorylation of phosphatidylinositol-4-phosphate by the type 1 phosphatidylinositol phosphate (PIP) kinases²⁹. PIP kinase type 1γ (PIPK1γ) was proposed to be the main enzyme responsible

for PtdIns(4,5)P₂ synthesis at synapses³⁰. PIPK1γ is also present at focal adhesion sites, where it participates in a feedback loop controlling adhesion mechanisms^{31,32}. It is a cytosolic protein whose recruitment to the plasma membrane is mediated by the interaction with small GTPases and with talin^{30–34}. Here we have used a genetic approach to test the hypothesis that PtdIns(4,5)P₂ synthesis by PIPK1γ is involved in pre-synaptic function, with an emphasis on synaptic vesicle trafficking.

Postnatal lethality and abnormal PtdIns(4,5)P₂ metabolism

Exons 2–6 of the PIPK1γ gene (*Pip5k1c*; hereafter referred to as *PIPK1γ*), which encode most of the catalytic region, were deleted by homologous recombination in mouse (Fig. 1a), as confirmed by Southern blot analysis of embryonic stem (ES) cells (data not shown) and F₂ offspring (Fig. 1b). Lack of *PIPK1γ* expression was verified by immunoblotting of brain extracts from postnatal day 0 (P0) animals (Fig. 1d and Supplementary Fig. 1). The genotypes of the F₂ offspring followed a mendelian distribution (data not shown) and all knockout animals died postnatally within 24 h. They were recognizable soon after birth by their impaired motility and by the lack of milk in their stomachs (Fig. 1d), but no obvious anatomical anomalies were observed. No difference in the expression level of a variety of synaptic proteins analysed or of the PIPK1γ interactors Arf6 and talin was found (Fig. 1d and not shown).

Levels of phosphatidylinositol, PIP and PtdInsP₂ in neonatal brain were analysed by electrospray ionization mass spectrometry (ESI–MS)³⁵. In this method, each of these phospholipid species appears as multiple peaks reflecting different fatty acid composition, with the C_{38:4} form (38 carbons and 4 double bonds accounted for by stearic and arachidonic acid, respectively) being by far the most abundant³⁵. Inspection of MS profiles revealed a 40% reduction in PtdInsP₂ levels in the brain of *PIPK1γ*^{−/−} animals, affecting the C_{38:4} peak (Fig. 2a) as well as other less abundant peaks (data not shown). In contrast, the levels of phosphatidylinositol and PIP appeared to be unchanged. ESI–MS does not allow discrimination between PtdInsP₂ isomers; however, most of the PtdInsP₂ in the brain is represented by the PtdIns(4,5)P₂

isomer (>95%), thus indicating the major role of *PIP1 γ* in mediating PtdIns(4,5)P₂ synthesis in the brain.

We also investigated whether the balance between PtdInsP₂ synthesizing and catabolizing activity was altered in the brain cytosol of *PIP1 γ ^{-/-}* animals. Inositol phospholipid-containing liposomes were incubated in the presence of wild-type or *PIP1 γ ^{-/-}* brain cytosol and [γ ³²P]ATP, and then analysed by thin layer chromatography (TLC). A substantially lower amount of radio-labelled PtdInsP₂, but no change in the labelling of phosphatidic acid and PIP, was observed in the *PIP1 γ ^{-/-}* samples (Fig. 2b). For comparison, incubation with brain cytosol from synaptojanin 1 knockout mice resulted in the opposite change; that is, an increase in PtdInsP₂ labelling (Fig. 2b)⁸.

PtdIns(4,5)P₂ synthesis in brain extracts was previously shown to be stimulated by GTP- γ S, a slowly hydrolysable analogue of GTP³⁶. This effect had been attributed, at least in part, to the small GTPase Arf6 on PIPK1 γ ³³. Accordingly, GTP- γ S stimulated the accumulation of [γ ³²P]-labelled PtdInsP₂ observed for control brain cytosol, but had little effect on PtdInsP₂ synthesis in *PIP1 γ ^{-/-}* cytosol (Fig. 2c).

Depolarization of synapses and synaptosomes produces a burst of phospholipase-C-mediated PtdIns(4,5)P₂ hydrolysis³⁷. It also induces PtdIns(4,5)P₂ synthesis, as revealed by metabolic labelling of synaptosomes with [γ ³²P]orthophosphate (not shown). The role of PIPK1 γ in this response was examined in metabolically active cell fragments from neonatal brains, obtained by the subcellular fractionation scheme typically used to prepare mature synaptosomes. TLC analysis of the labelled phospholipids revealed both a significant decrease in the labelling of PtdIns(4,5)P₂ in *PIP1 γ ^{-/-}* membranes (Fig. 2d) and a complete absence of the depolarization-induced increase in PtdInsP₂ synthesis (Fig. 2e). An analysis of the evoked release of neurotransmitter could not be performed on the material derived from the neonatal brains. However, studies of synaptosomes from adult *PIP1 γ* heterozygous animals revealed a reduction of depolarization-dependent glutamate release relative to controls (Supplementary Fig. 2), thus suggesting that a defect

in basal and evoked PtdIns(4,5)P₂ synthesis has an impact on neurotransmitter secretion.

We also generated synaptojanin 1/PIPK1 γ double knockout mice. The lack of synaptojanin 1 did not compensate for that of PIPK1 γ at the organismal level, as these mice were indistinguishable from *PIP1 γ ^{-/-}* mice, did not feed and died within 24 h (data not shown). These results are consistent with distinct sites of action of PIPK1 γ and synaptojanin 1 in the regulation of the synaptic vesicle cycle (see below). As expected, however, incubation of brain cytosol from these mice with liposomes and [γ ³²P]ATP revealed that the absence of synaptojanin 1 abrogated the defect in the accumulation of labelled PtdInsP₂ produced by the PIPK1 γ deficiency (Fig. 2f) in this *in vitro* assay, in which PIPK1 γ and synaptojanin 1 are not spatially segregated.

Synaptic transmission defects

We used primary cultures of cortical neurons to examine the impact of reduced PtdIns(4,5)P₂ levels and impaired stimulation-dependent PtdIns(4,5)P₂ synthesis on synaptic transmission and functional synaptic vesicle pools. *PIP1 γ ^{-/-}* neurons differentiated *in vitro* similarly to controls based on the analysis of axonal or dendritic processes and of pre- and post-synaptic compartments using specific immunocytochemical markers (data not shown). Previous studies of synaptojanin 1 knockout neurons revealed defects primarily during high-frequency stimulation^{8,38}. Similar experiments were performed on *PIP1 γ ^{-/-}* neurons and focused, as was the case for synaptojanin 1 (ref. 38), on inhibitory synapses. Paired patch-clamp recordings were carried out on monosynaptically connected cells. No significant differences in the basic properties of evoked inhibitory post-synaptic currents (IPSCs) elicited by single pre-synaptic action potentials (mean amplitude, 10–90% rise time and 10–90% decay time) were observed between control and mutant (*PIP1 γ ^{-/-}*) neurons (data not shown). Paired-pulse depression (PPD) of IPSCs in response to two action potentials (inter-stimulus intervals = 25–250 ms) was also unchanged (Fig. 3a). Furthermore, analysis of the amplitude distributions of

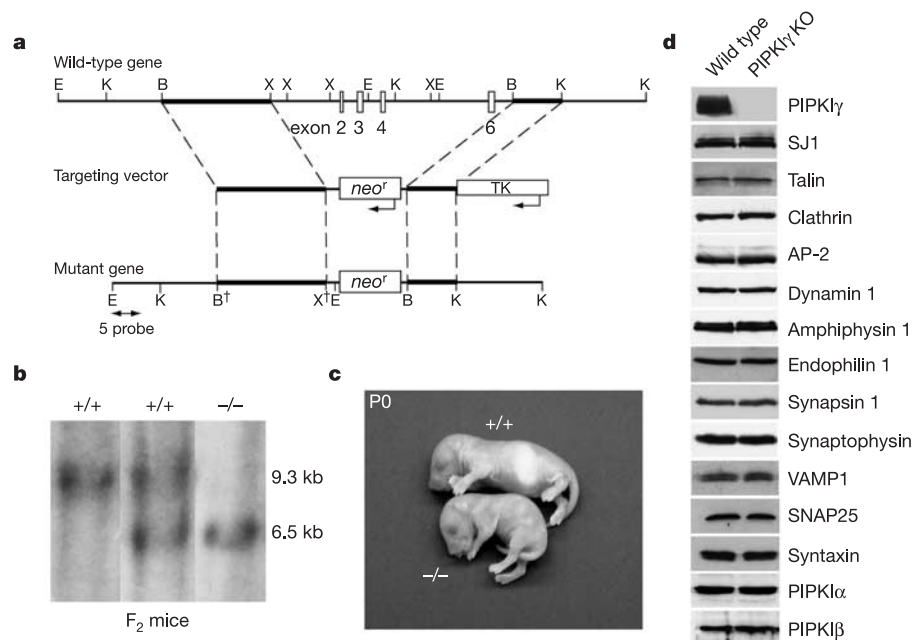


Figure 1 Early postnatal lethality in *PIP1 γ* -deficient mice. **a**, Targeting construct. *neo^R*, neomycin-resistance gene; TK, thymidine kinase gene; B, E, K, X denote the restriction enzymes *Bam*HI, *Eco*RI, *Kpn*I and *Xba*I, respectively; † indicates the loss of restriction site after recombination. **b**, Southern blot of *Eco*RI-digested genomic DNA from F₂

offspring showing homologous recombination. **c**, Newborn animals 10–15 h after birth. Note lack of milk in the *PIP1 γ ^{-/-}* pup. **d**, Western blots of brain extracts from P0 wild-type and knockout animals with a panel of antibodies.

miniature IPSCs (mIPSCs) resulting from the spontaneous release of individual synaptic vesicles did not reveal a significant difference in post-synaptic sensitivity between genotypes (mean amplitude = 24.5 ± 3.0 pA and 21.5 ± 2.8 pA for control and mutant synapses, respectively) (Fig. 3b). However, a significant

reduction (Kolmogorov–Smirnov test, $P < 0.0001$) in the frequency of mIPSCs was observed in mutant neurons (mean inter-event interval = $4,410.3 \pm 838.1$ ms) compared with control neurons ($2,695.0 \pm 336$ ms), consistent with a reduction in pre-synaptic function (Fig. 3b).

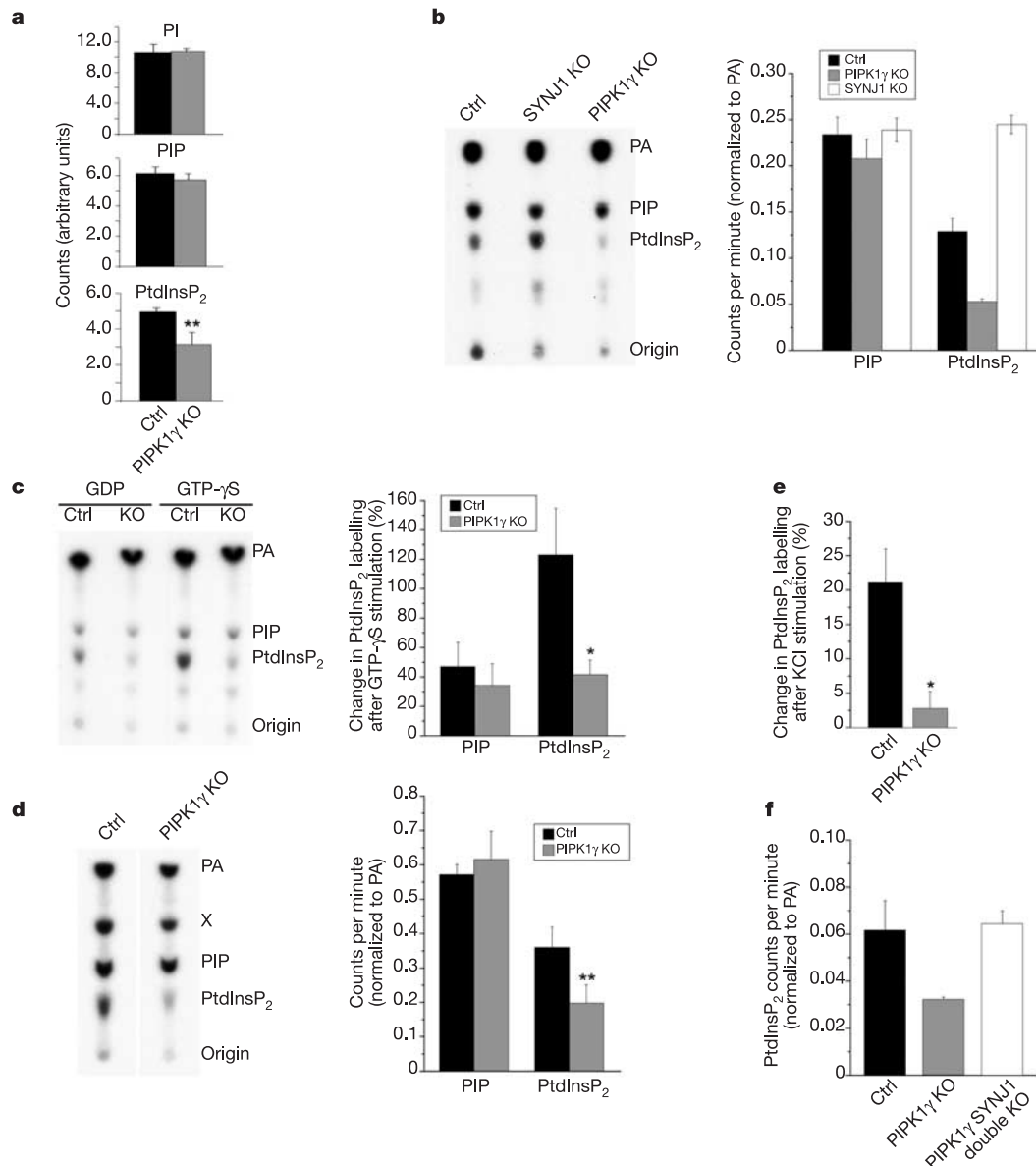


Figure 2 Alteration of PtdInsP₂ metabolism in *PIP1γ*^{-/-} mice. **a**, Levels of C_{38:4} phosphatidylinositol (PI), PIP and PtdInsP₂ in the brain of newborn *PIP1γ*^{+/+} (Ctrl) and *PIP1γ*^{-/-} (*PIP1γ* KO) mice as determined by ESI-MS ($n = 5$). Values are mean $\pm$ s.d.; asterisks indicate significant differences in a t -test ($P < 0.01$). **b**, TLC analysis of liposomes (Folch fraction) incubated for 10 min at 37 °C in the presence of [γ -³²P]ATP and brain cytosol from wild-type (Ctrl), *PIP1γ*^{-/-} (*PIP1γ* KO) and synaptojanin-1-deficient (SYNJ1 KO) animals. Left panel: autoradiography. PA, phosphatidic acid. Right panel: quantification of the radioactivity in the PIP and PtdInsP₂ spots obtained by phosphorimaging on duplicate TLCs from two independent experiments. Results are expressed as mean percentage of the total radioactivity in the TLC lane. Error bars are $\pm$ s.d. **c**, Reduced GTP-γS-stimulated PtdInsP₂ synthesis by *PIP1γ*^{-/-} brain cytosol. Cytosols were pre-incubated either with 1 mM GDP (or GDP-βS) or with 0.2 mM GTP-γS for 5 min at 37 °C before the experiment described in **b**. A representative autoradiography is shown on the left, whereas the right panel shows the quantification from three experiments. Values are mean $\pm$ s.d.; asterisk indicates significant differences in a t -test ($P < 0.05$). **d**, TLC analysis of phospholipids extracted

from brain P2 subcellular fractions metabolically labelled (30 min) with [³²P]orthophosphate. Left, autoradiography; right, quantification of the radioactivity on the PIP and PtdInsP₂ spots expressed as a percentage of the total radioactivity in the TLC lane. Values are mean $\pm$ s.d. ($n = 5$). Asterisks indicate significant differences in a t -test ($P < 0.01$). The lipid 'X' probably represents lysophosphatidic acid. **e**, Impaired depolarization-induced increase in PtdInsP₂ synthesis in P2 fractions of *PIP1γ*^{-/-} mice. P2 fractions, pre-labelled as in **d**, were stimulated with a high K⁺ concentration (47.5 mM) for 30 s at 37 °C or kept in normal buffer as a control. Columns show the percentage change in PtdInsP₂ labelling induced by high K⁺ treatment ($n = 5$ for control samples, $n = 4$ for *PIP1γ*^{-/-} samples); values are mean $\pm$ s.e.m.; asterisk indicates significant differences in an ANOVA test ($P < 0.05$). **f**, The PtdInsP₂ synthesis defect in *PIP1γ*^{-/-} brain cytosol can be rescued by a lack of synaptojanin 1. Brain cytosol from the indicated genotypes were incubated with liposomes and [γ -³²P]ATP for 10 min and then lipids were analysed and quantified by TLC ($n = 2$ for each genotype). Values are mean $\pm$ s.d.

Additional synaptic transmission defects were observed during high-frequency stimulation. As previously demonstrated³⁸, a train of 1,000 stimuli delivered at 20 Hz produces a rapid depression of the response that is followed by a nearly steady state where further depression occurs only very slowly. The early rapid depression (Fig. 3c, d) was significantly enhanced in *PIPK1 γ* ^{-/-} neurons ($n = 12$; $24.3 \pm 2.1\%$ of baseline after 100 pulses) relative to controls ($n = 38$; $33.9 \pm 2.3\%$ of baseline after 100 pulses; analysis of variance (ANOVA); $P < 0.001$; the difference becomes significant

after the third pulse of the train; $P < 0.05$, t -test), whereas at later times in the stimulation both types of neuron exhibited similar levels of depression ($12.0 \pm 2.1\%$ of baseline in control pairs and $14.8 \pm 1.5\%$ in mutant pairs after 1,000 pulses). At these late stages, a $\text{PtdIns}(4,5)\text{P}_2$ -dependent step may no longer be rate-limiting. Alternatively, compensatory mechanisms may occur. The enhancement of the initial depression was not seen at lower stimulation frequencies (2 Hz or 10 Hz (data not shown)).

The faster initial synaptic depression might reflect a smaller

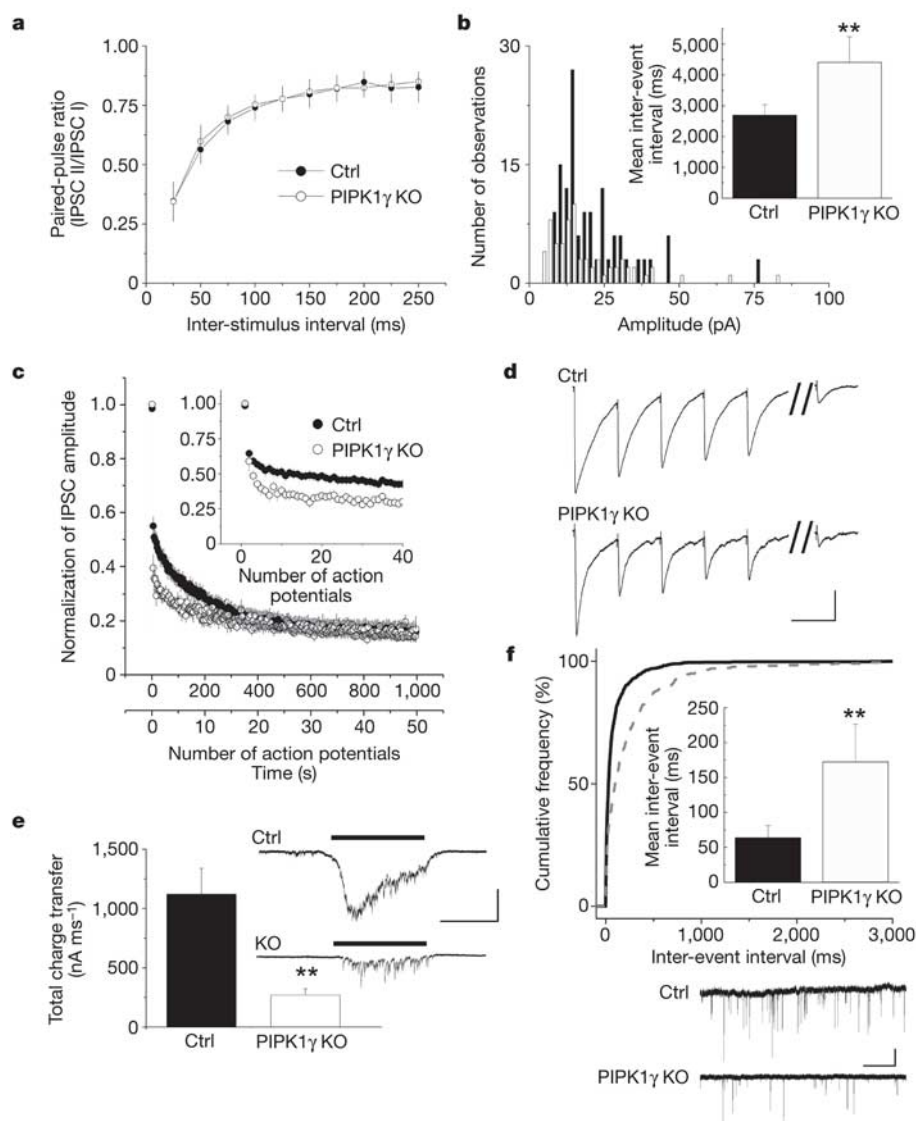


Figure 3 Synaptic transmission in *PIPK1 γ* ^{-/-} neurons grown in culture. **a**, PPD of IPSCs in control versus *PIPK1 γ* ^{-/-} neurons. IPSCs of mutant neurons did not differ significantly (ANOVA, $P = 0.89$) in response to a range of paired stimuli delivered at various inter-stimulus intervals (25–250 ms; $n = 5$ pairs). **b**, The distribution of mIPSC amplitudes does not differ between neurons of control and *PIPK1 γ* ^{-/-} mice (mean amplitude 24.5 ± 3.0 and 21.5 ± 2.8 pA, respectively; t -test, $P = 0.48$; Ctrl $n = 3$ cells; KO $n = 3$ cells). The frequency of mIPSCs is lower in mutant neurons, as reflected by a larger mean inter-event interval in the graph (inset). **c**, Time course of synaptic depression in response to 1,000 pre-synaptic action potentials delivered at 20 Hz. The rapid decline in IPSC amplitudes is steeper in *PIPK1 γ* ^{-/-} neurons ($n = 12$) than in control neurons ($n = 20$), whereas a similar nearly steady-state depression is reached by both neurons at later stages. Inset: responses to the first 40 action potentials of the train with an expanded x-axis (ANOVA, $P < 0.001$). **d**, Representative traces of the first through to the fifth and 1,000th IPSC in control and *PIPK1 γ* ^{-/-} cultures. Scale bar, 200 pA and 50 ms. **e**, Size of

the RRP of synaptic vesicles released in response to a 3-s pulse of hypertonic solution measured as the total charge transfer (corresponding to the total area under the current resulting from application of the hypertonic solution). Note the depression of the pool in knockout ($n = 18$ cells) compared with control neurons ($n = 21$ cells) (t -test, $P = 0.001$). Inset: representative traces obtained during application of the hypertonic solution (thick bars). Scale bar, 200 pA, 2 s. **f**, Decreased mEPSC frequency in *PIPK1 γ* ^{-/-} excitatory neurons. The cumulative frequency of the inter-event intervals is shown (top). The frequency of mEPSCs is lower in mutant neurons, as reflected by a larger mean inter-event interval in the graph (inset). The lower panel shows representative traces of mEPSCs (scale bar, 50 pA and 1 s). Spontaneously occurring mEPSCs were recorded from wild type ($n = 10$ cells, 5 cultures) and *PIPK1 γ* ^{-/-} ($n = 10$ cells, 3 cultures) neurons in the presence of 1 μM TTX and 10 μM bicuculline methiodide. For panels **b**, **e**, **f**, asterisks indicate statistical differences ($P < 0.01$). Error bars in this figure are $\pm$ s.e.m.

readily releasable pool (RRP) of vesicles, which represents the fusion-competent population of synaptic vesicles. In cultured neurons, this pool can be released in a Ca^{2+} -independent fashion in response to local application of a hypertonic solution^{39–41}. A fast-flow system was used to apply brief (3 s) pulses of hypertonic sucrose (500 mosM) to the soma and dendrites of recorded neurons. Measurements of the total charge transfer indicated that mutant inhibitory synapses had much smaller (t -test, $P < 0.001$) RRP ($(0.27 \pm 0.05) \times 10^6 \text{ pA ms}^{-1}$; $n = 18$) than control synapses ($(1.12 \pm 0.22) \times 10^6 \text{ pA ms}^{-1}$; $n = 21$) (Fig. 3e, f). Preliminary results revealed a decreased response in this assay also at mutant excitatory synapses (data not shown). Furthermore, consistent with results on inhibitory neurons, a significant reduction in the frequency of spontaneous miniature excitatory post-synaptic currents (mEPSCs) was observed (Fig. 3f). The mean inter-event interval was

increased from $64.0 \pm 17.2 \text{ ms}$ in controls to $172.7 \pm 53.6 \text{ ms}$ in $\text{PIP1}\gamma^{-/-}$ excitatory neurons (Kolmogorov–Smirnov test, $P < 0.001$). No difference was found in the mean mEPSC amplitude ($-35.4 \pm 8.2 \text{ pA}$ and $-33.1 \pm 3.4 \text{ pA}$ for control and mutant neurons, respectively). These findings indicate that defects are not limited to inhibitory synapses, as further supported by the reduced glutamate release from mutant synaptosomes (see above).

Smaller recycling pool and delayed recycling

Potential defects in synaptic vesicle trafficking were investigated by monitoring the kinetics of uptake and release of the fluorescent membrane tracer FM4-64 in cortical cultures. First, the recycling pool size was measured⁴². Cultures were loaded with FM4-64 using a variable number (20–600) of action potentials at 10 Hz. After a 10-min wash, the internalized dye was released by a train of 1,200

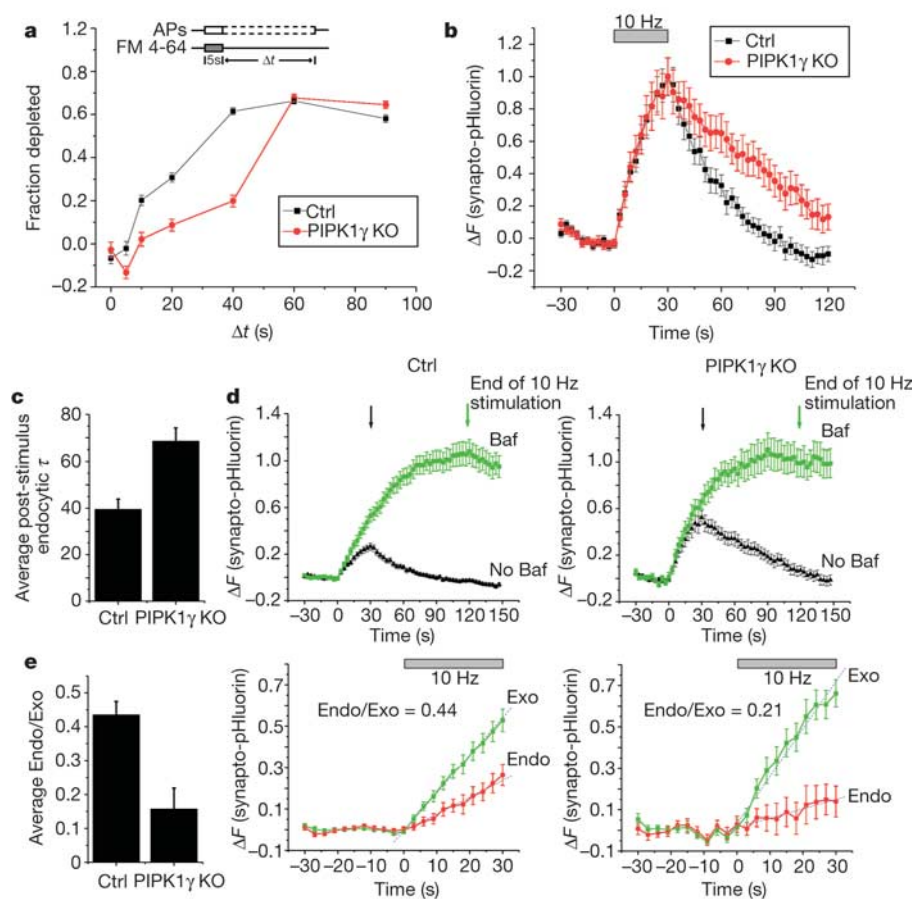


Figure 4 Slower rates of endocytosis and recycling in $\text{PIP1}\gamma^{-/-}$ neurons. **a**, Vesicle ‘re-priming’ time course. Inset: experimental protocol. Neuronal cultures were incubated with FM4-64 for 5 s while being stimulated at 20 Hz. The 20 Hz train exceeded the period of dye incubation by the time Δt . The residual internalized dye was quantified by a de-staining protocol after 10 min. The dye uptake at $\Delta t = 0 \text{ s}$ (F_0) denotes the total number of releasable vesicles mobilized by the stimulus. F_0 was calculated from the average value of bracketing runs of $\Delta t = 0 \text{ s}$. The dye uptake for all points, where $\Delta t > 0 \text{ s}$ ($F_{\Delta t}$), characterizes those vesicles that have regained fusion competence and released their dye content during the subsequent action potential stimulation. The fraction depleted is: $F_0 - F_{\Delta t} / F_0$. Each point was normalized to the averaged value of the bracketing runs of $\Delta t = 0 \text{ s}$. The data are from three separate experiments, in which 50–100 boutons were analysed in each case. Error bars are $\pm$ s.e.m. **b**, Time course of average fluorescence intensity over boutons expressing synapto-pHluorin, following stimulation with a train of 300 action potentials at 10 Hz. Values are mean $\pm$ s.e.m. **c**, Average τ_{decay} derived from multiple experiments from control and $\text{PIP1}\gamma^{-/-}$

cultures, following stimulation with a train of 300 action potentials at 10 Hz. Each curve was normalized to the peak value of ΔF (synapto-pHluorin) in each trace. Error bars are $\pm$ s.e.m. **d**, Control and mutant cultures were stimulated in the presence or absence of bafilomycin. In the absence of bafilomycin, the fluorescence signal reflects the net balance of exocytosis and endocytosis. In the presence of bafilomycin, exocytic events are trapped in an alkaline state and the fluorescence signal reflects exocytosis exclusively. Endocytosis during the stimulus was derived by subtracting the fluorescence trace in the absence of bafilomycin from the trace in the presence of bafilomycin ($\Delta F_{\text{endo}} = \Delta F_{\text{exo}} - \Delta F_{\text{exo-endo}}$). The traces obtained in the presence and in the absence of bafilomycin were normalized to the maximal stable fluorescent signal in the bafilomycin trace (31 and 20 boutons for control and mutant cultures, respectively). Error bars are $\pm$ s.e.m. **e**, Graph showing the average speed of endocytosis relative to exocytosis for wild type ($n = 15$) and knockout ($n = 16$) cultures (25–40 boutons per culture). Error bars are $\pm$ s.e.m.

action potentials at 10 Hz. As expected⁴², the amount of released dye, which reflects the recycling vesicle pool, increased for both genotypes as a function of the duration of the initial staining step (Supplementary Fig. 3). However, an overall 20% decrease in the pool size was observed in mutant nerve terminals.

Second, the recycling time (also referred to as 're-priming' time⁴²;

that is, the time that elapses before dye-loaded endocytic vesicles become available for a new round of exocytosis during continuous stimulation) was measured. Synapses were labelled with FM4-64 for 5 s by a 20 Hz stimulation that was maintained for a variable time, Δt , after the wash-out of the dye (inset in Fig. 4a). During this additional time, some internalized dye (F_0) starts being released as a result of recycling. After a 10-min rest, the total internalized and releasable fluorescence left in the terminal was determined by the application of a stimulus (see Methods) that produces complete destaining ($F_{\Delta t}$). The ratio $(F_0 - F_{\Delta t})/F_0$ denotes the fraction of the fluorescence that was released during the additional period of stimulation and reflects the gain in exocytosis competence of recently endocytosed vesicles. Synapses from both genotypes reached a similar maximal release of the recently internalized dye (nearly 67%) within the next 60 s of stimulation. However, mutant nerve terminals were slower in regaining release competence ($t_{1/2} \approx 20$ s and 45 s for control and mutant neurons, respectively), indicating a delay in membrane recycling.

Slower endocytosis kinetics

To determine whether the endocytic reaction itself was affected, a 'synapto-pHluorin'-based assay, which enables real-time measurement of synaptic vesicle exo- and endocytosis, was used^{43,44}. Synapto-pHluorin comprises a pH-sensitive variant of the green fluorescent protein (superecliptic pHluorin) fused to the carboxy terminus (and therefore exposed to the vesicle lumen) of the synaptic vesicle membrane protein synaptobrevin 2 (also known as VAMP2)⁴³. When expressed in neurons, most pHluorin resides in the acidic synaptic vesicle lumen where its fluorescence is low. After exocytosis, pHluorin becomes exposed to the neutral extracellular environment leading to an increase in fluorescence. Endocytosis and rapid re-acidification of the newly formed vesicles reverses this change⁴⁴. Neurons transfected with synapto-pHluorin were stimulated with 300 action potentials at 10 Hz (Fig. 4b, c). Although the action potential train increased fluorescence in nerve terminals from both genotypes, the recovery from the peak fluorescence (normalized to 1 in Fig. 4c) was 50% slower in mutant nerve terminals ($t_{1/2} = 39.4 \pm 4.4$ s for $PIPK1\gamma^{+/+}$ compared with 68.5 ± 5.7 s for $PIPK1\gamma^{-/-}$). Acid quenching experiments verified that re-acidification rates in both genotypes occur much faster than endocytosis (Supplementary Fig. 4).

The increase in synapto-pHluorin fluorescence during stimulation reflects the net balance of exocytosis and endocytosis. To assess further the potential endocytic defects during repetitive stimulation, the assay was performed in the presence of bafilomycin A1 (ref. 45). This drug prevents the acidification of newly endocytosed vesicles by blocking their proton pump⁴⁵. Thus, the fluorescence signal varies only as a function of exocytic rate, which was found to be comparable between the two genotypes (0.94 ± 0.07 for $PIPK1\gamma^{+/+}$ and 1.0 ± 0.09 for $PIPK1\gamma^{-/-}$ synapses, respectively; values are arbitrary fluorescence units). A comparison of fluorescence increase during stimulation in the presence and absence of bafilomycin A1 allows for the indirect measurement of endocytic rate⁴⁵ and therefore for the determination of an endocytosis/exocytosis ratio (Fig. 4d, e). The average ratio was smaller in mutant synapses (0.16 ± 0.06) than in controls (0.44 ± 0.04), indicating a slowing of endocytosis relative to exocytosis during stimulation.

Impairment of clathrin-mediated endocytosis

As PtdIns(4,5)P₂ is thought to participate in the recruitment of clathrin adaptors and some accessory factors of clathrin, the delay in endocytosis may be explained at least partially by an impairment in clathrin coat dynamics. A defect in clathrin-mediated budding might also account for the delay revealed by the 'recycling/re-priming assay' (Fig. 4a), because during stimulation clathrin-coated vesicles are thought to bud in parallel both from the plasma membrane and from endosome-like structures that result from

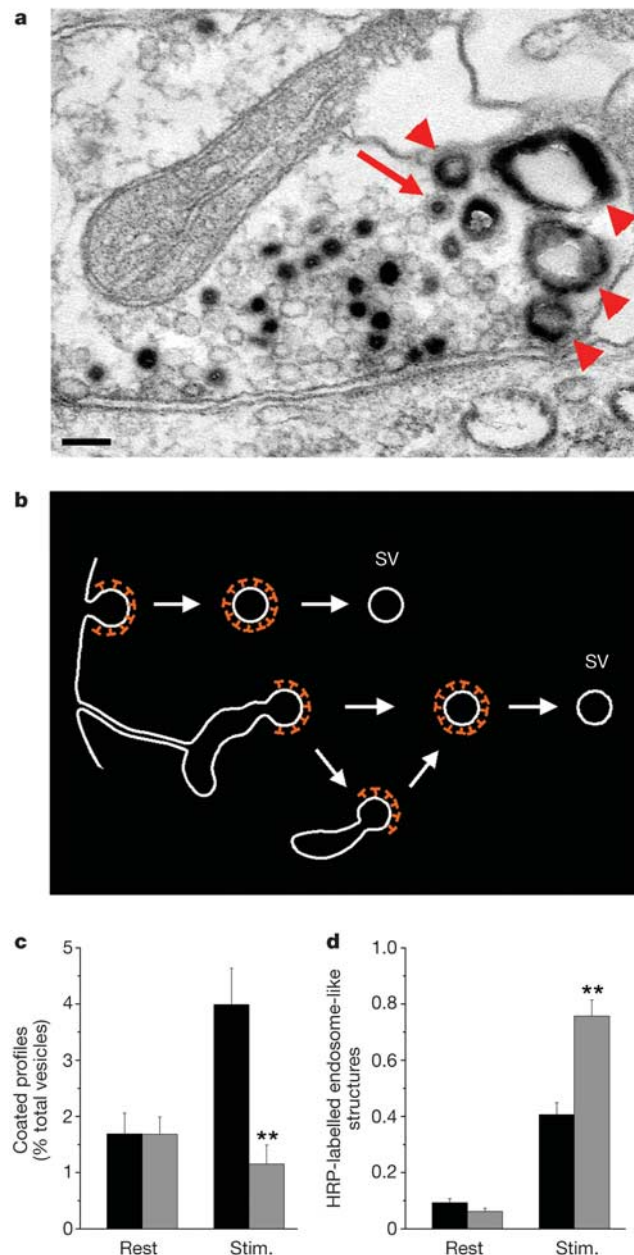


Figure 5 Defects in membrane recycling in $PIPK1\gamma$ -deficient nerve terminals as revealed by electron microscopy. **a**, Electron micrograph of a nerve terminal stimulated with a high K^+ concentration in the presence of HRP. Arrowheads point to endosome-like structures; arrows indicate a clathrin-coated vesicle. Scale bar, 100 nm. **b**, Schematic cartoon of clathrin-mediated endocytosis at the synapse (see also ref. 16). SV, synaptic vesicle. **c**, The percentage of coated vesicular profiles (relative to all vesicles) in nerve endings of cultured control and $PIPK1\gamma^{-/-}$ cortical neurons is the same at rest, but is not increased by stimulation (90 s with high K^+) in the mutant neurons. **d**, Stimulation with high K^+ induces a larger increase in HRP-labelled endosome-like endocytic intermediates in $PIPK1\gamma^{-/-}$ neurons than in control neurons. See Methods for the quantification. Double asterisks indicate statistical significance of knockout relative to control in the stimulated condition ($P < 0.01$, ANOVA). Error bars in **c** and **d** are $\pm$ s.e.m.

bulk endocytosis¹⁶ (Fig. 5a). An electron microscopy analysis of cultured neurons labelled with the endocytic tracer horseradish peroxidase (HRP) was performed. Briefly, neurons were incubated for 90 s in control medium and subsequently either stimulated with 90 mM KCl for 90 s or kept in control medium for the same time, always in the presence of HRP, before fixation (Fig. 5a). No obvious differences between control and mutant neurons were observed in the absence of stimulation (Fig. 5c, d). However, two major differences were observed after stimulation. First, the percentage of coated vesicles relative to the total number of vesicles (coated vesicles + synaptic vesicles) was 3.5-fold lower in *PIPK1 γ* ^{-/-} nerve terminals (1.15 ± 0.34) than in control nerve terminals (3.99 ± 0.65), and was virtually unchanged from the resting level (Fig. 5c). Second, HRP-labelled endosome-like structures were more abundant in mutant neurons. Their membrane profiles were increased by 85% after stimulation relative to controls (Fig. 5d). These observations strongly suggest a defect in the clathrin-dependent reformation of synaptic vesicles and are consistent with a back-up of endocytic traffic both in the plasma membrane and in endosome-like structures (see also Fig. 5b).

General considerations

This study demonstrates the principal role of PIPK1 γ in the regulation of PtdIns(4,5)P₂ metabolism in the brain, its importance for normal neuronal physiology and its essential function for postnatal life. Surprisingly, no major developmental defects were observed in *PIPK1 γ* ^{-/-} mice, despite evidence of a role for this kinase at sites of focal adhesions in all cells^{30–32}. Although the precise cause of death of *PIPK1 γ* ^{-/-} mice remains undetermined, the inability of these mice to feed after birth strongly suggests the occurrence of major neurological and/or behavioural defects. Furthermore, functional analysis of *PIPK1 γ* ^{-/-} neurons in culture revealed defects in synaptic transmission that may result, at least in part, from an impairment of synaptic vesicle traffic at multiple steps of their cycle. We have revealed not only an endocytic delay, but also changes suggestive of exocytic defects, such as a smaller RRP, reduced miniature current frequency and enhanced rapid depression during prolonged high-frequency stimulation. These exocytic defects might be accounted for, in part, by a delay in the time required for the replacement of docked and primed vesicles that have undergone exocytosis with new fusion-competent vesicles. This is a particularly attractive possibility because a role for PtdIns(4,5)P₂ in priming was inferred from studies of secretory granule secretion in permeabilized neuroendocrine cells²⁶, and the PtdIns(4,5)P₂ metabolite DAG was shown to be involved in synaptic vesicle priming via its binding to Munc13 (ref. 12). Interactions of PtdIns(4,5)P₂ with synaptotagmin and with additional proteins of the exocytic machinery may also be involved⁴⁶. As Arf6 is a positive regulator of PIPK1 γ ^{33,34}, it is of interest that Munc13 interacts with an Arf6–GEF, mSec7-1 (also known as cytohesin-2)⁴⁷, whose recruitment to the plasma membrane is in turn regulated by phosphoinositides⁴. Thus, a positive feedback mechanism aimed at controlling PtdIns(4,5)P₂ levels may have an important role at synapses and may be disrupted in *PIPK1 γ* ^{-/-} mice.

Our findings point to a dual role of pre-synaptic PtdIns(4,5)P₂ in exo- and in endocytosis, and demonstrate an important function of PIPK1 γ in maintaining appropriate levels of the synaptic pool of PtdIns(4,5)P₂, both in resting conditions and after stimulation, when PtdIns(4,5)P₂ consumption by phospholipase C, synaptotagmin 1 and possibly other enzymes, such as phosphatidylinositol-3-OH kinases, is increased. Our results support the occurrence of a cycle of PtdIns(4,5)P₂ synthesis and hydrolysis nested within the synaptic vesicle cycle in which PIPK1 γ is important both for exocytosis and for endocytosis, whereas synaptotagmin 1 is important for clathrin uncoating. They do not rule out other actions mediated by the regulatory role of PtdIns(4,5)P₂ and its metabolites (including inositol trisphosphate, DAG, fatty acids and other phospho-

inositides) on ion channels, transporters, cytoskeletal proteins and other effectors^{1,2}. In addition, whereas we have focused our investigations on pre-synaptic function, additional effects of PIPK1 γ disruption on neuronal function, and on postsynaptic function in particular, can be anticipated. □

Methods

Generation of *PIPK1 γ* knockout mice

A genomic library from 129SVJ mice was screened with mouse *PIPK1 γ* complementary DNA to generate a targeting vector using pGT-N29 (NEB) (Fig. 1a). Part of the construct was generated by Incyte Genomics Inc. Exons 2–6 were replaced by the neomycin-resistance gene. The HSV-1 thymidine kinase gene was added at the end of the short arm. The vector was linearized and electroporated into ES cells, which were then selected according to standard procedures⁸. Southern blot screening was performed with a polymerase chain reaction (PCR)-generated 0.7-kilobase probe located upstream of the long arm (Fig. 1a). Positive clones were microinjected into blastocysts. The resulting chimaeras were bred with C57BL/6 mice. Germline transmission was achieved and heterozygous animals were identified by Southern blot analysis (Fig. 1b) and/or PCR.

Immunoblotting

Western blotting was performed by standard procedures⁸. All of the antibodies have been previously described^{30,31}, with the exception of rabbit anti-PIPK1 α and - β (a gift from J. Kunz).

Lipid biochemistry

ESI-MS analysis of phosphoinositides and phosphoinositide radiolabelling using liposomes, [³²P]ATP and brain cytosol were performed as described^{8,35}. Radioactivity was quantified using a phosphorimager. In some experiments, cytosols were pre-incubated for 5 min at 37 °C with 1 mM GDP or 0.2 mM GTP- γ S. Brain 'P2' subcellular fractions for metabolic labelling were prepared as follows: neonatal brain tissue was homogenized in 0.4 ml 25 mM HEPES, pH 7.4, 0.32 M sucrose (HB buffer) and centrifuged at 5,000 r.p.m. in a microfuge. The resulting supernatant was diluted in 4 ml HB buffer and centrifuged at 11,500 r.p.m. for 15 min at 4 °C in an SS34 Sorvall centrifuge rotor. The pellet was re-suspended in 4 ml 130 mM NaCl, 5 mM KCl, 2 mM CaCl₂, 2 mM MgCl₂, 25 mM HEPES (pH 7.33), 30 mM glucose (control buffer) followed by re-centrifugation and re-suspension in the same buffer at a final protein concentration of 2 mg ml⁻¹. 100- μ l aliquots were incubated with [³²P]orthophosphate at 2 mCi ml⁻¹ for 30 min at 37 °C. At the end of this incubation, samples were subjected to lipid extraction. Counts in the PtdIns(4,5)P₂ spot were normalized to those in the phosphatidic acid spot, which were not affected by the various genotypes. Lipids were extracted by chloroform/methanol (2:1) supplemented with 20 mg ml⁻¹ cold phosphoinositides (Sigma) and processed for TLC.

Experiments on cultured neurons

Cultures were prepared as described from the cortex of newborn animals^{8,42} and grown for 2–3 weeks before the experiments.

Electrophysiology

Paired whole-cell patch clamp recordings were performed at 34–35 °C with Axopatch 200B patch clamp amplifiers (Axon Instruments) as described⁴⁸. For the PPD experiments, the paired-pulse ratio (amplitude of the second IPSC divided by the amplitude of the first IPSC) was plotted for each inter-stimulus interval (average of three trials for each interval). For the time course of depression experiment, the peak amplitudes of each IPSC in the train were normalized to the average baseline amplitude (average of 5–10 consecutive IPSCs evoked at 0.05 Hz before delivery of the train). During the train, the IPSC amplitudes were normalized to the preceding baseline values obtained with a 0.2 Hz stimulation. For RRP measurements, 3-s pulses of hypertonic (500 mosM) sucrose were applied near neuronal perikarya using a fast-flow system (Warner Instruments) in the presence of tetrodotoxin (TTX, 3 μ M). The total charge transfer (corresponding to the total area under the current) was analysed using the pClamp 9.0 program. Miniature IPSCs/EPSCs were analysed from the traces obtained in the presence of TTX. All experiments were performed blind to the genotype of the cultures and the data are presented as mean $\pm$ s.e.m. Statistical significance was assessed using unpaired *t*-tests or one-way ANOVA as appropriate.

Imaging experiments

FM4-64 experiments were performed using confocal laser scanning microscopy and electrical field stimulation, as previously described⁴². The protocol used to obtain a maximal dye de-staining was 900 action potentials at 10 Hz followed by a 2-min rest period and a second train of 1,200 action potentials at 20 Hz. Synapto-pHluorin experiments were performed according to published procedures^{44,45}.

Electron microscopy

Neurons were labelled with HRP (Sigma) at 10 mg ml⁻¹ for 90 s at 37 °C in control buffer (Tyrode buffer) containing 10 μ M CNQX (6-cyano-7-nitroquinoxaline-2,3-dione) and 50 μ M AP5(D(-)-2-amino-5-phosphonopentanoic acid) (Research Biochemicals). Neurons were then either kept in control buffer for another 90 s or stimulated for 90 s with 90 mM KCl in the continued presence of HRP and then hypotonically fixed in 1% glutaraldehyde, 50 mM sodium cacodylate buffer (pH 7.4) for 1 h at room temperature to allow for the visualization of clathrin coats¹⁶. Morphometry was performed under blind experimental

conditions by analysing 2,000–3,500 objects (synaptic vesicles, coated vesicles and HRP-labelled endosome-like structures) at a magnification of $\times 63,000$ in approximately 40 synapses for each condition. Coated vesicles were defined as structures with a diameter <100 nm exhibiting the bristle-like coat characteristics of clathrin. For each synapse, the perimeter of HRP-labelled 'endosome-like' structures and the outer perimeter of the nerve terminal ('plasma membrane') were measured with NIH Image. In Fig. 5d, the perimeter of endosome-like structures was normalized to that of the plasma membrane. Statistics were performed using ANOVA. Results obtained from two independent experiments were pooled. Results obtained with neurons from *PIPK1 γ ^{+/+}* mice were comparable to those obtained with neurons from *PIPK1 γ ^{-/-}* mice and were consequently pooled and referred to as 'control' cultures.

Received 5 June; accepted 27 July 2004; doi:10.1038/nature02896.

- De Camilli, P., Emr, S. D., McPherson, P. S. & Novick, P. Phosphoinositides as regulators in membrane traffic. *Science* **271**, 1533–1539 (1996).
- Hilgemann, D. W., Feng, S. & Nasuhoglu, C. The complex and intriguing lives of PIP₂ with ion channels and transporters. *Sci. STKE* **2001**, RE19 (2001).
- Martin, T. F. PI(4,5)P(2) regulation of surface membrane traffic. *Curr. Opin. Cell Biol.* **13**, 493–499 (2001).
- Czech, M. P. Dynamics of phosphoinositides in membrane retrieval and insertion. *Annu. Rev. Physiol.* **65**, 791–815 (2003).
- Valtorta, F. & Meldolesi, J. The presynaptic compartment: signals and targets. *Semin. Cell Biol.* **5**, 211–219 (1994).
- Wang, S. S. & Augustine, G. J. Confocal imaging and local photolysis of caged compounds: dual probes of synaptic function. *Neuron* **15**, 755–760 (1995).
- Wiedemann, C., Schafer, T., Burger, M. M. & Sihra, T. S. An essential role for a small synaptic vesicle-associated phosphatidylinositol 4-kinase in neurotransmitter release. *J. Neurosci.* **18**, 5594–5602 (1998).
- Cremona, O. *et al.* Essential role of phosphoinositide metabolism in synaptic vesicle recycling. *Cell* **99**, 179–188 (1999).
- Lackner, M. R., Nurrish, S. J. & Kaplan, J. M. Facilitation of synaptic transmission by EGL-30 Gq α and EGL-4 PLC β : DAG binding to UNC-13 is required to stimulate acetylcholine release. *Neuron* **24**, 335–346 (1999).
- Gad, H. *et al.* Fission and uncoating of synaptic clathrin-coated vesicles are perturbed by disruption of interactions with the SH3 domain of endophilin. *Neuron* **27**, 301–312 (2000).
- Harris, T. W., Hartwig, E., Horvitz, H. R. & Jorgensen, E. M. Mutations in synaptotagmin disrupt synaptic vesicle recycling. *J. Cell Biol.* **150**, 589–600 (2000).
- Rhee, J. S. *et al.* Beta phorbol ester- and diacylglycerol-induced augmentation of transmitter release is mediated by Munc13s and not by PKCs. *Cell* **108**, 121–133 (2002).
- Micheva, K. D., Buchanan, J., Holz, R. W. & Smith, S. J. Retrograde regulation of synaptic vesicle endocytosis and recycling. *Nature Neurosci.* **6**, 925–932 (2003).
- Murthy, V. N. & De Camilli, P. Cell biology of the presynaptic terminal. *Annu. Rev. Neurosci.* **26**, 701–728 (2003).
- Heuser, J. E. & Reese, T. S. Evidence for recycling of synaptic vesicle membrane during transmitter release at the frog neuromuscular junction. *J. Cell Biol.* **57**, 315–344 (1973).
- Takei, K., Mundigl, O., Daniell, L. & De Camilli, P. The synaptic vesicle cycle: A single vesicle budding step involving clathrin and dynamin. *J. Cell Biol.* **133**, 1237–1250 (1996).
- Fesce, R. & Meldolesi, J. Peeping at the vesicle kiss. *Nature Cell Biol.* **1**, E3–E4 (1999).
- Aravanis, A. M., Pyle, J. L. & Tsien, R. W. Single synaptic vesicles fusing transiently and successively without loss of identity. *Nature* **423**, 643–647 (2003).
- Gandhi, S. P. & Stevens, C. F. Three modes of synaptic vesicular recycling revealed by single-vesicle imaging. *Nature* **423**, 607–613 (2003).
- McPherson, P. S. *et al.* A presynaptic inositol-5-phosphatase. *Nature* **379**, 353–357 (1996).
- Kim, W. T. *et al.* Delayed reentry of recycling vesicles into the fusion-competent synaptic vesicle pool in synaptotagmin 1 knockout mice. *Proc. Natl Acad. Sci. USA* **99**, 17143–17148 (2002).
- Verstreken, P. *et al.* Synaptotagmin is recruited by endophilin to promote synaptic vesicle uncoating. *Neuron* **40**, 733–748 (2003).
- Takei, K. *et al.* Generation of coated intermediates of clathrin-mediated endocytosis on protein-free liposomes. *Cell* **94**, 131–141 (1998).
- Ford, M. G. *et al.* Curvature of clathrin-coated pits driven by epsin. *Nature* **419**, 361–366 (2002).
- Evans, P. R. & Owen, D. J. Endocytosis and vesicle trafficking. *Curr. Opin. Struct. Biol.* **12**, 814–821 (2002).
- Hay, J. C. *et al.* ATP-dependent inositolide phosphorylation required for Ca(2+)-activated secretion. *Nature* **374**, 173–177 (1995).
- Holz, R. W. *et al.* A pleckstrin homology domain specific for phosphatidylinositol-4,5-bisphosphate

- (PtdIns-4,5-P₂) and fused to green fluorescent protein identifies plasma membrane PtdIns-4,5-P₂ as being important in exocytosis. *J. Biol. Chem.* **275**, 17878–17885 (2000).
- Khvotchev, M. & Sudhof, T. C. Newly synthesized phosphatidylinositol phosphates are required for synaptic norepinephrine but not glutamate or γ -aminobutyric acid (GABA) release. *J. Biol. Chem.* **273**, 21451–21454 (1998).
- Doughman, R. L., Firestone, A. J. & Anderson, R. A. Phosphatidylinositol phosphate kinases put PI4,5P(2) in its place. *J. Membr. Biol.* **194**, 77–89 (2003).
- Wenk, M. R. *et al.* PIP kinase I γ is the major PI(4,5)P(2) synthesizing enzyme at the synapse. *Neuron* **32**, 79–88 (2001).
- Di Paolo, G. *et al.* Recruitment and regulation of phosphatidylinositol phosphate kinase type I γ by the FERM domain of talin. *Nature* **420**, 85–89 (2002).
- Ling, K., Doughman, R. L., Firestone, A. J., Bunce, M. W. & Anderson, R. A. Type I γ phosphatidylinositol phosphate kinase targets and regulates focal adhesions. *Nature* **420**, 89–93 (2002).
- Krauss, M. *et al.* ARF6 stimulates clathrin/AP-2 recruitment to synaptic membranes by activating phosphatidylinositol phosphate kinase type I γ . *J. Cell Biol.* **162**, 113–124 (2003).
- Aikawa, Y. & Martin, T. F. ARF6 regulates a plasma membrane pool of phosphatidylinositol(4,5) bisphosphate required for regulated exocytosis. *J. Cell Biol.* **162**, 647–659 (2003).
- Wenk, M. R. *et al.* Phosphoinositide profiling in complex lipid mixtures using electrospray ionization mass spectrometry. *Nature Biotechnol.* **21**, 813–817 (2003).
- Honda, A. *et al.* Phosphatidylinositol 4-phosphate 5-kinase α is a downstream effector of the small G protein ARF6 in membrane ruffle formation. *Cell* **99**, 521–532 (1999).
- Audigier, S. M., Wang, J. K. & Greengard, P. Membrane depolarization and carbamoylcholine stimulate phosphatidylinositol turnover in intact nerve terminals. *Proc. Natl Acad. Sci. USA* **85**, 2859–2863 (1988).
- Luthi, A. *et al.* Synaptotagmin 1 contributes to maintaining the stability of GABAergic transmission in primary cultures of cortical neurons. *J. Neurosci.* **21**, 9101–9111 (2001).
- Stevens, C. F. & Tsujimoto, T. Estimates for the pool size of releasable quanta at a single central synapse and for the time required to refill the pool. *Proc. Natl Acad. Sci. USA* **92**, 846–849 (1995).
- Rosenmund, C. & Stevens, C. F. Definition of the readily releasable pool of vesicles at hippocampal synapses. *Neuron* **16**, 1197–1207 (1996).
- Kashani, A. H., Chen, B. M. & Grinnell, A. D. Hypertonic enhancement of transmitter release from frog motor nerve terminals: Ca²⁺ independence and role of integrins. *J. Physiol. (Lond.)* **530**, 243–252 (2001).
- Ryan, T. A. Inhibitors of myosin light chain kinase block synaptic vesicle pool mobilization during action potential firing. *J. Neurosci.* **19**, 1317–1323 (1999).
- Miesenböck, G., De Angelis, D. A. & Rothman, J. E. Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. *Nature* **394**, 192–195 (1998).
- Sankaranarayanan, S. & Ryan, T. A. Real-time measurements of vesicle-SNARE recycling in synapses of the central nervous system. *Nature Cell Biol.* **2**, 197–204 (2000).
- Sankaranarayanan, S. & Ryan, T. A. Calcium accelerates endocytosis of vSNAREs at hippocampal synapses. *Nature Neurosci.* **4**, 129–136 (2001).
- Bai, J., Tucker, W. C. & Chapman, E. R. PIP(2) increases the speed of response of synaptotagmin and steers its membrane-penetration activity toward the plasma membrane. *Nature Struct. Mol. Biol.* **11**, 36–44 (2004).
- Neeb, A., Koch, H., Schurmann, A. & Brose, N. Direct interaction between the ARF-specific guanine nucleotide exchange factor msec7-1 and presynaptic Munc13-1. *Eur. J. Cell Biol.* **78**, 533–538 (1999).
- Fitzsimonds, R. M., Song, H. J. & Poo, M. M. Propagation of activity-dependent synaptic depression in simple neural networks. *Nature* **388**, 439–448 (1997).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We would like to thank L. Liu, L. Lucast, L. Daniell and W. Yan for technical assistance; J. Kunz and R. Jahn for the antibodies; and J. Morgan and O. Cremona for critical reading of the manuscript. This work was supported in part by National Institutes of Health grants to P.D.C., M.W., T.A.R. and R.M.F., and by pilot grants from the Yale Center for Genomic and Proteomics to P.D.C. and M.W.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.D.C. (pietro.decamilli@yale.edu).

Inhibition of carbonate synthesis in acidic oceans on early Mars

Alberto G. Fairén¹, David Fernández-Remolar², James M. Dohm³, Victor R. Baker^{3,4} & Ricardo Amils^{1,2}

¹Centro de Biología Molecular, CSIC–Universidad Autónoma de Madrid, 28049-Cantoblanco, Madrid, Spain

²Centro de Astrobiología (CSIC–INTA), 28850-Torrejón de Ardoz, Madrid, Spain

³Department of Hydrology and Water Resources, ⁴Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

Several lines of evidence have recently reinforced the hypothesis that an ocean existed on early Mars^{1–7}. Carbonates are accordingly expected to have formed from oceanic sedimentation of carbon dioxide from the ancient martian atmosphere^{7,8}. But spectral imaging of the martian surface has revealed the presence of only a small amount of carbonate, widely distributed in the martian dust⁹. Here we examine the feasibility of carbonate synthesis in ancient martian oceans using aqueous equilibrium calculations. We show that partial pressures of atmospheric carbon dioxide in the range 0.8–4 bar, in the presence of up to 13.5 mM sulphate and 0.8 mM iron in sea water⁸, result in an acidic oceanic environment with a pH of less than 6.2. This precludes the formation of siderite, usually expected to be the first major carbonate mineral to precipitate⁸. We conclude that extensive interaction between an atmosphere dominated by carbon dioxide and a lasting sulphate- and iron-enriched acidic ocean on early Mars is a plausible explanation for the observed absence of carbonates.

A survey of the literature suggests four alternative classical explanations for the detection of only small concentrations of carbonate minerals: (1) no primary carbonate formation in a cold and/or dry environment; (2) the inability of the Thermal Emission Spectrometer (TES; on board the Mars Global Surveyor spacecraft) and the Thermal Emission Imaging System (THEMIS; on board the Mars Odyssey spacecraft) to detect carbonates; (3) secondary chemical alteration of ancient carbonate sediments; or (4) carbonates that are obscured by younger rock materials. The first explanation, which is in disagreement with the geomorphological evidence, supports the interpretation of early martian conditions as having been similar to those prevailing at present¹⁰. The second highlights the possibility that substantial regional carbonate deposits can be 100% exposed and remain undetected at the sensitivity of TES/THEMIS¹¹. The third includes the possibility of water vapour and sulphates combining to form acid rain⁴ and thereby promoting chemical decomposition of superficial carbonate layers, acid-fog weathering¹², and/or photodecomposition¹³. And the fourth includes large carbonate deposits hidden beneath several-centimetres-thick secondary alteration rinds that envelop rock materials on the surface¹⁴, or carbonates masked by the formation of relatively recent volumetrically abundant soils¹⁴; it also includes other resurfacing processes that may have mantled carbonate deposits, such as eolian deposition during the dominant phases of Mars' mostly frozen state¹⁵, sedimentation in a later Hesperian-age ocean that occupied the northern plains⁵, emplacement of basaltic lavas and pyroclastic materials in and around the Tharsis and Elysium magmatic complexes³, and the deposition of 100-m-thick Vastitas Borealis Formation materials, a potentially ice-rich sublimation residue of outflow events¹⁶.

Using aqueous equilibrium calculations, we analyse here the feasibility of carbonate synthesis in ancient martian oceans, taking into account the geochemical cycles that were probably in operation then. To this end, our analysis is based on two key assumptions: (1)

that well over 90% of the carbonate rock formation on Earth occurs in open bodies of liquid water¹⁷; and (2) that the fate of the martian oceans was rapid freezing solid followed by sublimation and cold-trapping of ice at higher latitudes^{1,10}. Thus, the problem of carbonate formation on early Mars basically concerns the chemical conditions operating in open bodies of liquid water. Local sedimentation driven by sequential evaporation in closed basins⁸ is considered here an interesting, but minor, mode of water evolution on early Mars; and where such processes are likely to have occurred, the subsequent increase of iron and sulphate in the solution would have further lowered the pH, thus maintaining ionic undersaturation with respect to carbonates. On the other hand, more recent transitory reduced lakes would not have been able to control the process of carbonate synthesis either, as Mars has been mainly cold and dry throughout the Hesperian and the Amazonian, except during brief and intriguing episodes^{5,15}.

To raise the mean surface temperature of Mars above the freezing point, a greater atmospheric CO₂ pressure is required. The present CO₂ pressure (6 mbar), in addition to that adsorbed in the regolith (30–40 mbar; ref. 18) and the minimum quantity sequestered as ice or clathrate in the poles, represent nearly 40 mbar of CO₂. If the elimination processes by impact erosion and pick-up ion sputtering removed 95–99% of the Noachian atmosphere¹⁸, then the martian surface CO₂ pressures 4 Gyr ago could have been of the order of 800–4,000 mbar. However, without the additional warming provided by the presence of other strong greenhouse gasses, such as CH₄ and NH₃ (ref. 19), even 4 bar of CO₂ would have been insufficient to yield mean global temperatures above the freezing point. Carbon dioxide ice clouds in the troposphere may have also contributed to warming early Mars, as ice particles reflect the outgoing thermal infrared radiation back to the surface²⁰.

Nevertheless, this approach for high CO₂ pressures in early Mars has been classically disputed by two lines of arguments: the loss of atmospheric volatiles¹⁸, and the lack of carbonates on the martian surface¹⁰. But substantial amounts of atmospheric CO₂ may have been maintained during the Noachian by volcanic outgassing from Tharsis^{2,3,5}. In addition, if the oceans were acidic, the lack of carbonates no longer qualifies as a reliable indicator of CO₂ partial pressures in the Noachian martian atmosphere. Consequently, the possibility of substantial atmospheric CO₂ pressures in early Mars cannot be ruled out. Later, nearly all of the martian atmosphere would have been lost by impact erosion during the heavy bombardment period^{18,21}, and by sputtering acting throughout most of the martian history¹⁸.

Such a carbon dioxide-dominated atmosphere in early Mars would have rendered an ocean mildly acidic by releasing free protons in the sequence $\text{H}_2\text{O} + \text{CO}_2(\text{g}) \leftrightarrow \text{H}_2\text{CO}_3 \leftrightarrow \text{H}^+ + \text{HCO}_3^-$. The addition of iron contributed even more to oceanic acidification. Also, sufficient iron in solution acted as an excellent buffer, as appreciable amounts of iron can be dissolved in moderately acidic to neutral solutions. Hydrolysis of ferrous iron generates free protons by the equilibrium $\text{Fe}^{2+} + 2\text{H}_2\text{O} \leftrightarrow \text{Fe}(\text{OH})_2 + 2\text{H}^+$, and the process is much more effective if the iron is in the form of ferric iron: $\text{Fe}^{3+} + 3\text{H}_2\text{O} \leftrightarrow \text{Fe}(\text{OH})_3 + 3\text{H}^+$.

The potential source of iron in terrestrial Archaean oceans was iron emitted from hydrothermal vent fluids produced by intense submarine magmatic activity creating hot Fe²⁺-enriched plumes²²; and the subsequent photolytic oxidation of ferrous iron to Fe³⁺ would provide the ocean with its major oxidized species²³. Consequently, iron is thought to have been abundant in the oceans of early Earth, reaching concentrations as high as 50 μM (ref. 24). On early Mars, the iron-rich nature of the mantle and active volcanism³ produced igneous minerals containing ferrous iron. In fact, the martian surface is much richer in iron than the surface of Earth: the reddish colour of the martian surface is due to the optical properties of ferric-iron-bearing minerals present in the oxidized surface layer²⁵. In this sense, up to 800 μM of Fe²⁺ for basin feedwater⁸

on the surface of early Mars has been considered a reasonable estimate for an iron-rich fluid.

Moreover, jarosite minerals, as well as evaporite deposits containing magnesium sulphate salts, a typical sublimation residue, have been detected by the Mars Exploration Rover (MER) Opportunity in the sediments deposited at Meridiani Planum. The amount of salts detected, up to 40% in the outcrop⁷, leads us to expect that sulphate levels in ancient martian oceans were at least at the same order of magnitude as those derived from weathered ultramafic rocks on Earth (0.18 mM; ref. 8), since the salinity of present-day terrestrial oceanic water is 3.5×10^5 p.p.m. (ref. 26). Following the Opportunity results, however, we assume a Fe^{3+} -enriched solution in equilibrium with jarosite²⁷ at least locally and/or temporarily, so potentially displaying even higher sulphate levels, up to 13.5 mM (see Table 1).

For the Noachian oceans, aqueous thermodynamic calculations considering a solution enriched in iron hydroxides and sulphate result in a pH between 5.3 and 6.2 for the siderite- Fe^{2+} equilibrium; and when the ferrous iron was photolytically oxidized to ferric, the final pH was between 1.9 and 2.1 for the siderite- Fe^{3+} equilibrium (Fig. 1; see Methods). The subsequent evolution of the acidic environment over time results in Fe^{3+} -sulphatic species dominating the chemistry of the oceans (Fig. 2). Thus, the TES analyses, which indicate an absence of carbonate minerals at the surface of Mars, can be comprehensively explained by the acidity of the ancient martian oceans. These results address the recently posed contradiction between TES-detected martian mineralogy, and planetary geomorphology and MER-unveiled geochemistry, the former reporting small amounts of disseminated carbonates and the latter evidencing the geomorphologic modification and chemical alteration of surface materials through significant amounts of running water during long periods of early Mars history. Also, this new scenario is qualitatively different from all those argued before (and listed above), as they account for masking or altering the carbonates actually synthesized in a reducing environment. On the contrary, in our model, carbonate rock layers simply never formed, but instead we envisage the possibility of a warmer and wetter Mars, where extensive interaction between significant amounts of moderate to extreme acidic liquid water and a CO_2 atmosphere might have occurred (Fig. 3).

The viability of the model proposed here for ancient Mars is supported by the latest observational evidence provided by MER Opportunity at Meridiani Planum. Also on the modern Earth, the headwaters of the Tinto river system (in the southwest of Spain), an extreme acidic environment controlled by iron biogeochemistry, produces ferric-iron-enriched sediments dominated by sulphate and oxihydroxide parageneses, resulting in goethite, haematite and jarosite, analogous to the minerals found in Meridiani²⁸. In the living Tinto river system, the values of pH, redox potential, and

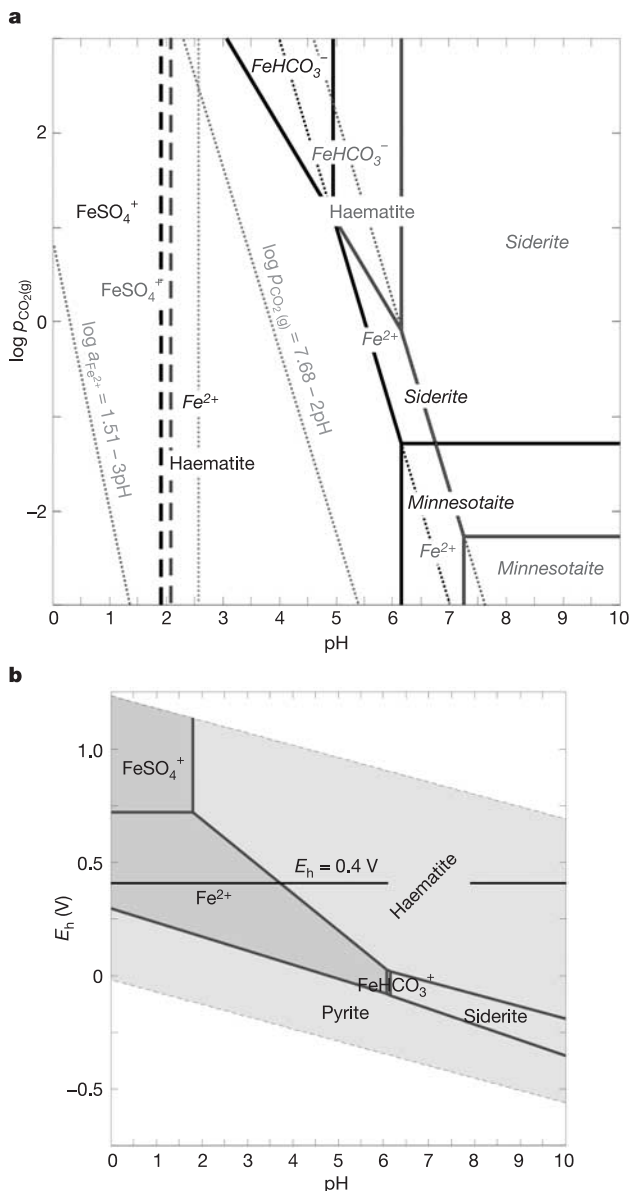


Figure 1 Stability boundaries of siderite and minnesotaite (under reducing conditions) and acidic solutions and haematite (under oxidizing conditions) on early Mars. **a**, Stability boundaries obtained as a function of CO_2 pressure and pH, calculated at 273 K for four different solutions: black lines represent a solution containing Fe and $\text{SiO}_2(\text{aq})$ at their upper limits (800 μM and 1 mM, respectively); grey lines represent the lower limits assumed here (50 μM and 180 μM); solid lines represent reducing conditions (Fe^{2+}) and standard $[\text{SO}_4^{2-}] = 0.18 \text{ mM}$; and dashed lines represent oxidizing conditions (Fe^{3+}) and upper and lower limits for $[\text{SO}_4^{2-}] = 6.5$ (black) and 13.5 (grey) mM, driving to the formation of haematite and acidic solutions from which jarosite is expected to form at $\text{pH} \approx 2$ (boundary of jarosite stability). Dotted lines represent, from left to right: (1) the pH buffer when Fe^{3+} controls the solution, under acidic and oxidizing conditions, assuming goethite as the ultimate chemical product ($\text{Fe}^{3+} + \text{H}_2\text{O} \leftrightarrow \text{FeOOH} (\text{goethite}) + 3\text{H}^+$); (2) the equilibrium pH of jarosite ($(\text{K}_2\text{Fe}_6(\text{SO}_4)_4(\text{OH})_{12}) + 6\text{H}^+ \leftrightarrow 2\text{SO}_4^{2-} + 6\text{H}_2\text{O} + \text{K}^+ + 3\text{Fe}^{3+}$) when minimum values for Fe^{3+} (50 mM) and sulphates (180 mM) are considered; (3) the changes in pH exclusively controlled by the p_{CO_2} ; and (4 and 5) the relationship between CO_2 pressure and pH for each modelled solution ($\text{pH} = 5.6 - 0.5 \log p_{\text{CO}_2}$ and $\text{pH} = 6.1 - 0.5 \log p_{\text{CO}_2}$, respectively). All constituents included in the solutions for displaying this graphic are quantified in Table 1. **b**, E_h -pH diagram assuming the lower values of Table 1, at $p_{\text{CO}_2} = 1 \text{ atm}$ and $T = 273 \text{ K}$. The stability fields of jarosite (acidic oxidizing solutions with FeSO_4^+), haematite and siderite are represented. Dark grey indicates phase solutions, while light grey indicates minerals. Jarosite formation is favoured by $E_h > 0.4 \text{ V}$ and $\text{pH} < 2$.

Table 1 Assumed average composition for surface waters on early Mars

Constituent	Concentration (p.p.m.)	Molality ($\times 10^{-3} \text{ mol kg}^{-1}$)	Activity
$\text{SiO}_2(\text{aq})$	10.8–60.1	0.18*–1.0	$10^{-3.75}$ – 10^{-3}
$\text{Fe}^{2+}/\text{Fe}^{3+}$	3–44.7	0.05†–0.8	$10^{-4.3}$ – $10^{-3.1}$
SO_4^{2-}	17.3 (624–1,263)	0.18 (6.5–13.5)‡	$10^{-3.745}$ ($10^{-2.19}$ – $10^{-1.88}$)
Mg^{2+}	24.3	1.0	10^{-3}
Cl^-	23.0	0.65	10^{-7}
Ca^{2+}	20.0	0.5	$10^{-3.19}$
Na^{2+}	18.4	0.8	$10^{-3.1}$
K^+	2.7	0.07	$10^{-4.16}$

Concentration and molality data after ref. 8, except as shown by footnote symbols as follows.

*From ref. 30.

†From ref. 24.

‡Obtained assuming a solution at 273 K in equilibrium with jarosite for $a_{\text{Fe}^{3+}} = 10^{-4.3}$ and $a_{\text{Fe}^{2+}} = 10^{-3.1}$, respectively, as governed by the equation $\log a_{\text{SO}_4^{2-}} = 3 \log a_{\text{H}^+} - 0.5 \log a_{\text{K}^+} - 1.5 \log a_{\text{Fe}^{3+}} - 2.91$.

HCO_3^- is not a fixed parameter here, as it has been swapped for $\text{CO}_2(\text{g})$, the y-axis variable in Fig. 1a.

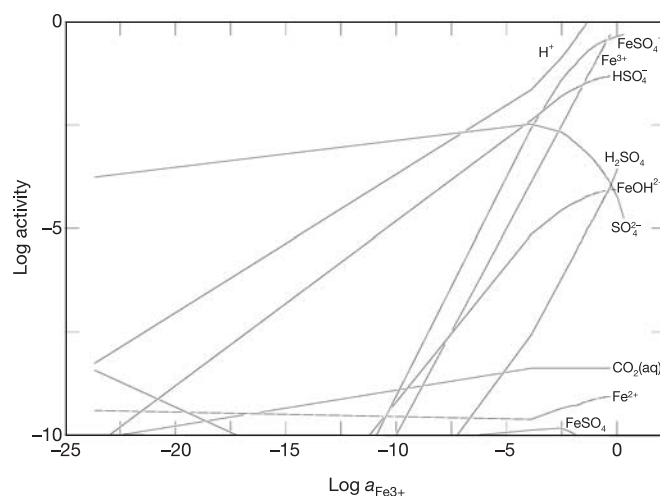


Figure 2 Evolution of the acidic martian environment over time. Evolution of solution and gas compositions, starting from an initial pH = 7 and $E_h = 0.4$ V, at increasing concentration of Fe^{3+} and H_2SO_4 . Fe^{3+} concentration ranges from 1.4×10^{-5} M to 0.5 M, a current record in modern acidic waters at the Tinto river²⁸. Although the ferric ion can acidify different solutions when added at high concentrations, sulphuric solutions generated by H_2S and SO_2 photo-oxidation may indeed generate acidic conditions at first, acting as an acidic matrix for later reactions that produce jarositic associations after

evaporation. Thus, the system has been coupled to an increased concentration of H_2SO_4 from the initial equilibrium to 10^{-4} M, a rough estimation for the acidic waters of the Tinto river²⁸. Although a high diversity of chemical species emerged, following the MER Opportunity results only S and Fe species are displayed, that is, those that would dominate the hypothetical chemistry of the early martian oceans. In this case, Fe^{3+} –sulphatic species would favour the ferric sulphate precipitation.

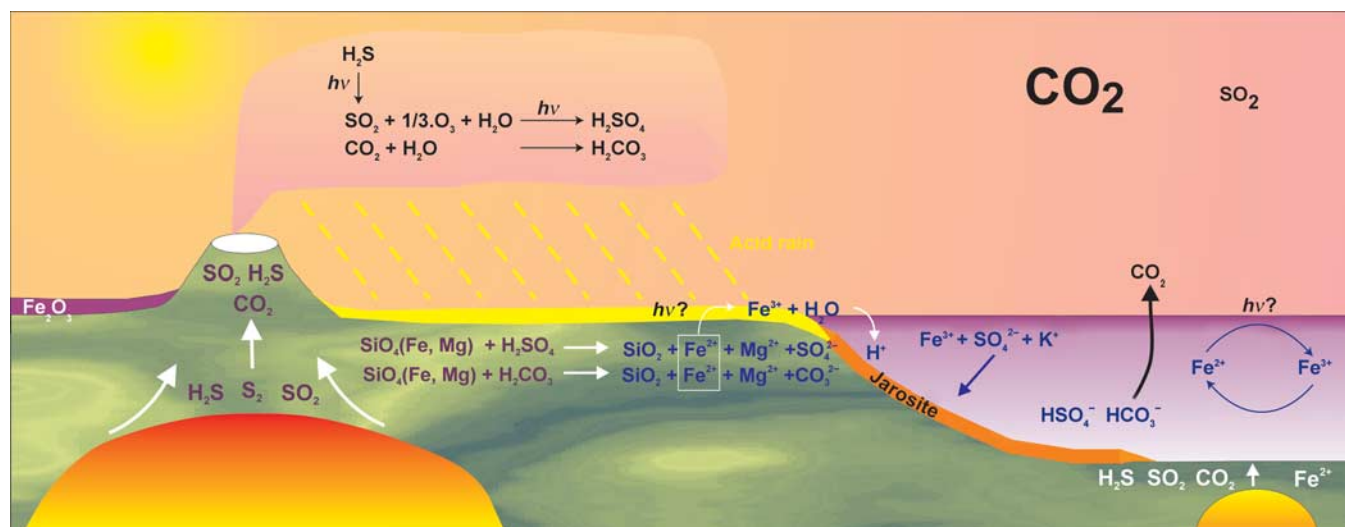


Figure 3 Schematic representation of atmosphere–land–ocean interactions generating acidic environments in early Mars. The acidic attack of ultramafic crust is driven by H_2SO_4 (generated by volcanic SO_2 photolysis) and H_2CO_3 , and consequently weathering of basalt releases Fe and Mg to the sea water. As ferrous iron oxidizes to ferric, Fe^{3+} drives the water more and more acidic, as also does SO_2 forming dilute sulphuric acid. From these solutions, jarosite as well as Mg sulphates are expected to precipitate, as highlighted by

MER Opportunity at Meridiani. Also, the acidic weathering process would generate a sulphate layer over the basaltic crust, potentially able to evolve to haematite by dehydration of the ferric sulphates and related iron-bearing minerals. Under such conditions, HCO_3^- , the main source of carbonates, is at very low concentration, and therefore all C is expected to be in the form of $\text{CO}_2(\text{aq})$ and $\text{CO}_2(\text{g})$, with higher residence time in the atmosphere and therefore maintaining long-term milder surface temperatures.

Fe^{3+} and sulphate concentration remain constant in different years under variable rain regimes. Life is highly diverse in the Tinto system, allowing us to suggest comparable ancient acidic aquatic habitats hosting a putative early biosphere on Mars. In fact, if biological inhabitation of early Mars is considered plausible, moderate acidic oceans represent the closest terrestrial analogue for a biogenic environment, similar to that where life originated on the Hadean–Early Archaean Earth^{23,26}. In the Tinto river system, the remobilization of iron and its subsequent precipitation as ferric

minerals preserves biological remains sometimes in remarkable detail, suggesting an analogous record in the rock exposures derived from the acidic martian early environments. □

Methods

We use the values listed in Table 1 to estimate limits on the pH and composition of the putative Noachian ocean, which includes considering the addition of supplementary cations such as Mg^{2+} , Ca^{2+} , K^+ or Na^+ , and silica (a by-product of the acid dissolution of ferromagnesian silicates, typical of martian igneous rocks in such acidic oceans), as they are effective solutes that raise the pH (ref. 8).

Sedimentation models in depositional environments establish that siderite (FeCO_3) is

always the first major carbonate to precipitate from ferriferous solutions, because it is the most insoluble of the major common carbonates; whereas minnesotaite ($\text{Fe}_3\text{Si}_4\text{O}_{10}(\text{OH})_2$), a phyllosilicate compositionally close to greenalite, is the simplest of the possible ferrous silicates, as well as a stable phase in terrestrial sediments. According to the reactivity of siderite and minnesotaite given by their corresponding dissociation constants²⁹, the equations governing the limits of the stability fields for both minerals with respect to oceanic pH are $\text{pH} = 3.97 - 0.5 \log p_{\text{CO}_2} - 0.5 \log a_{\text{Fe}^{2+}}$ and $\text{pH} = 5.1 - 0.5 \log a_{\text{Fe}^{2+}}$, respectively, for amorphous precipitates.

The combination of pH, redox potential and concentration of CO_2 determines the stability of carbonates as possible phases, and their precipitation by inorganic processes depends on the saturation state of the solution with respect to these minerals. Oceanic pH specifically related to an equilibrium between aqueous iron and siderite can be determined from the relation $\log K = \log a_{\text{Fe}^{2+}} + \log p_{\text{CO}_2} - 2 \log a_{\text{H}^+}$, as calculated in Fig. 1a. Using values at 273 K and a mean concentration for Fe^{2+} in sea water of between 50 and 800 μM , a pH between 5.3 and 6.2 for the siderite– Fe^{2+} equilibrium is obtained for an ocean that occurs in assumed atmospheric CO_2 pressures of between 0.8 and 4 bar, contrasting to the pH value in modern oceans on Earth, about 8.1; when iron finally photooxidized to Fe^{3+} , pH varies between 1.9 and 2.1. Assuming silica concentrations close to the saturation point of oceanic waters on Earth (180 μM , ref. 30) or even higher for water on early Mars (1 mM, ref. 8), and assuming that no Earth-like biological recycling occurred on Mars during the Noachian period, these relations define the boundaries for siderite, minnesotaite, haematite and jarosite precipitation shown in Fig. 1. System evolution over time, coupled to an increased concentration of H_2SO_4 with Fe^{3+} , which would favour the precipitation of ferric sulphate, is shown in Fig. 2. For the conservative constraints assumed here (see Table 1), neither precipitation nor the metastability of siderite are noted. Increasing the assumed values will result in even more acidic conditions, thus making carbonate formation even less likely. Following the established model for the Earth, additional powerful contributors to the acidity of the Noachian oceans may have been volcanic sulphur-rich gases causing acid rain, as indicated in Fig. 3.

Received 7 January; accepted 23 July 2004; doi:10.1038/nature02911.

- Clifford, S. M. & Parker, T. J. The evolution of the martian hydrosphere: Implications for the fate of a primordial ocean and the current state of the northern plains. *Icarus* **154**, 40–79 (2001).
- Phillips, R. J. *et al.* Ancient geodynamics and global-scale hydrology on Mars. *Science* **291**, 2587–2591 (2001).
- Dohm, J. M. *et al.* Ancient drainage basin of the Tharsis region, Mars: Potential source for outflow channel systems and putative oceans or paleolakes. *J. Geophys. Res.* **106**, 32943–32958 (2001).
- Craddock, R. A. & Howard, A. D. The case for rainfall on a warm, wet early Mars. *J. Geophys. Res.* **107**, doi:10.1029/2001JE001505 (2002).
- Fairén, A. G. *et al.* Episodic flood inundations of the northern plains of Mars. *Icarus* **165**, 53–67 (2003).
- Malin, M. C. & Edgett, K. S. Evidence for persistent flow and aqueous sedimentation on early Mars. *Science* **302**, 1931–1934 (2003).
- Moore, J. M. Blueberry fields for ever. *Nature* **428**, 711–712 (2004).
- Catling, D. C. A chemical model for evaporites on early Mars: Possible sedimentary tracers of the early climate and implications for exploration. *J. Geophys. Res.* **104**, 16453–16469 (1999).
- Bandfield, J. L., Glotch, T. D. & Christensen, P. R. Spectroscopic identification of carbonate minerals in the martian dust. *Science* **301**, 1084–1087 (2003).
- Carr, M. H. & Head, J. W. Oceans on Mars: An assessment of the observational evidence and possible fate. *J. Geophys. Res.* **108**, doi:10.1029/2002JE001963 (2003).
- Kirkland, L. E., Herr, K. C. & Adams, P. M. Infrared stealthy surfaces: Why TES and THEMIS may miss some substantial mineral deposits on Mars and implications for remote sensing of planetary surfaces. *J. Geophys. Res.* **108**, doi:10.1029/2003JE002105 (2003).
- Huguenin, R. L. J. The formation of goethite and hydrated clay minerals on Mars. *J. Geophys. Res.* **79**, 3895–3905 (1974).
- Mukhin, L. M., Koscheev, A. P., Dikov, Yu. P., Huth, J. & Wänke, H. Experimental simulations of the photo-decomposition of carbonates and sulphates on Mars. *Nature* **379**, 141–143 (1996).
- Clark, B. C. On the non-observability of carbonates on Mars. *5th Mars Conf. Abstr.* 6214 (Lunar and Planetary Institute, Houston, Texas, 1999).
- Baker, V. R. Water and the Martian landscape. *Nature* **412**, 228–236 (2001).
- Head, J. W. III, Kreslavsky, M. A. & Pratt, S. Northern lowlands of Mars: Evidence for wide-spread volcanic flooding and tectonic deformation in the Hesperian Period. *J. Geophys. Res.* **107**, doi:10.1029/2000JE001445 (2002).
- Bhattacharyya, A. & Friedman, G. M. *Modern Carbonate Environments* (Hutchinson Ross, Stroudsburg, Pennsylvania, 1983).
- Brain, D. A. & Jakosky, B. M. Atmospheric loss since the onset of the martian geologic record: Combined role of impact erosion and sputtering. *J. Geophys. Res.* **103**, 22689–22694 (1998).
- Squyres, S. W. & Kasting, J. F. Early Mars: How warm and how wet? *Science* **265**, 744–749 (1994).
- Forget, F. & Pierrehumbert, R. T. Warming early Mars with carbon dioxide clouds that scatter infrared radiation. *Science* **278**, 1273–1276 (1997).
- Melosh, H. J. & Vickery, A. M. Impact erosion of the primordial atmosphere of Mars. *Nature* **338**, 487–489 (1989).
- Barley, M. E., Pickard, A. L. & Sylvestre, P. J. Emplacement of a large igneous province as possible cause of banded iron formation 2.45 billion years ago. *Nature* **385**, 55–58 (1997).
- Russell, M. J. & Hall, A. J. The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *J. Geol. Soc.* **154**, 377–402 (1997).
- Holland, H. D. The oceans: A possible source of iron in iron-formations. *Econ. Geol.* **68**, 1169–1172 (1973).
- Zuber, M. T. The crust and mantle of Mars. *Nature* **412**, 220–227 (2001).
- Schaefer, M. W. Aqueous geochemistry on early Mars. *Geochim. Cosmochim. Acta* **57**, 4619–4625 (1993).
- Burns, R. G. Ferric sulfates on Mars. *J. Geophys. Res.* **92**, 570–574 (1987).
- Fernández-Remolar, D. *et al.* The Tinto river, an extreme acidic environment under control of iron, as an analog of the *Terra Meridiani* hematite site of Mars. *Planet. Space Sci.* **52**, 239–248 (2003).

- Johnson, J. W., Oelkers, E. H. & Helgeson, H. C. SUPCRT92: A software package for calculating the standard molal thermodynamic properties of minerals, gases, aqueous species, and reactions from 1 to 5000 bars and 0° to 1000 °C. (Earth Sciences Department, Lawrence Livermore Laboratory, 1991).
- Bruland, K. W. in *Chemical Oceanography* 8 (eds Riley, J. P. & Chester, R.) 157–220 (Academic, London, 1983).

Acknowledgements Special acknowledgements to the MER team, as their compelling evidence probing the acidity of martian palaeoenvironments was unfolded while our work was in progress, and resulted in adjustments in our model following our initial submitted draft. We also thank S. Clifford, I. Fairchild and J. Kasting for comments and suggestions that refined and focused this paper.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.G.F. (agfairén@cbm.uam.es).

A high-intensity highly coherent soft X-ray femtosecond laser seeded by a high harmonic beam

Ph. Zeitoun¹, G. Faivre¹, S. Sebban¹, T. Mocek¹, A. Hallou¹, M. Fajardo², D. Aubert³, Ph. Balcou¹, F. Burgy¹, D. Douillet¹, S. Kazamias⁴, G. de Lachèze-Murel³, T. Lefrou¹, S. le Pape⁵, P. Mercère¹, H. Merdji⁶, A. S. Morlens¹, J. P. Rousseau¹ & C. Valentin¹

¹Laboratoire d'Optique Appliquée, chemin de la Humière, 91761 Palaiseau, France

²Centro de Física dos Plasmas, Instituto Superior Técnico, Avenida Rovisco Pais, 1049-001 Lisboa, Portugal

³CEA/DIF, BP 12, 91680 Bruyères-le-châtel, France

⁴Laboratoire d'Interaction du Rayonnement X avec la Matière, bâtiment 350, Université Paris-Sud, 91405 Orsay, France

⁵Laboratoire pour l'Utilisation des Lasers Intenses, École Polytechnique, 91128 Palaiseau, France

⁶CEA/SPAM, bâtiment 522, Centre d'Études de Saclay, 91191 Gif-sur-Yvette, France

Synchrotrons have for decades provided invaluable sources of soft X-rays, the application of which has led to significant progress in many areas of science and technology. But future applications of soft X-rays—in structural biology, for example—anticipate the need for pulses with much shorter duration (femtoseconds) and much higher energy (millijoules) than those delivered by synchrotrons. Soft X-ray free-electron lasers¹ should fulfil these requirements but will be limited in number; the pressure on beamtime is therefore likely to be considerable. Laser-driven soft X-ray sources offer a comparatively inexpensive and widely available alternative, but have encountered practical bottlenecks in the quest for high intensities. Here we establish and characterize a soft X-ray laser chain that shows how these bottlenecks can in principle be overcome. By combining the high optical quality available from high-harmonic laser sources (as a seed beam) with a highly energetic soft X-ray laser plasma amplifier, we produce a tabletop soft X-ray femtosecond laser operating at 10 Hz and exhibiting full saturation, high energy, high coherence and full polarization. This technique should be readily applicable on all existing laser-driven soft X-ray facilities.

A laser chain consists of several stages: a seed oscillator, delivering a pulse of perfect optical properties, followed by amplifiers, set to increase beam energy. We studied the extension of this scheme to the soft X-ray range. The main difficulties while designing an amplification chain are first to produce a seed with high optical quality, then to maintain the quality over the amplification stages and finally to reduce the spurious self-emission of the amplifiers to a

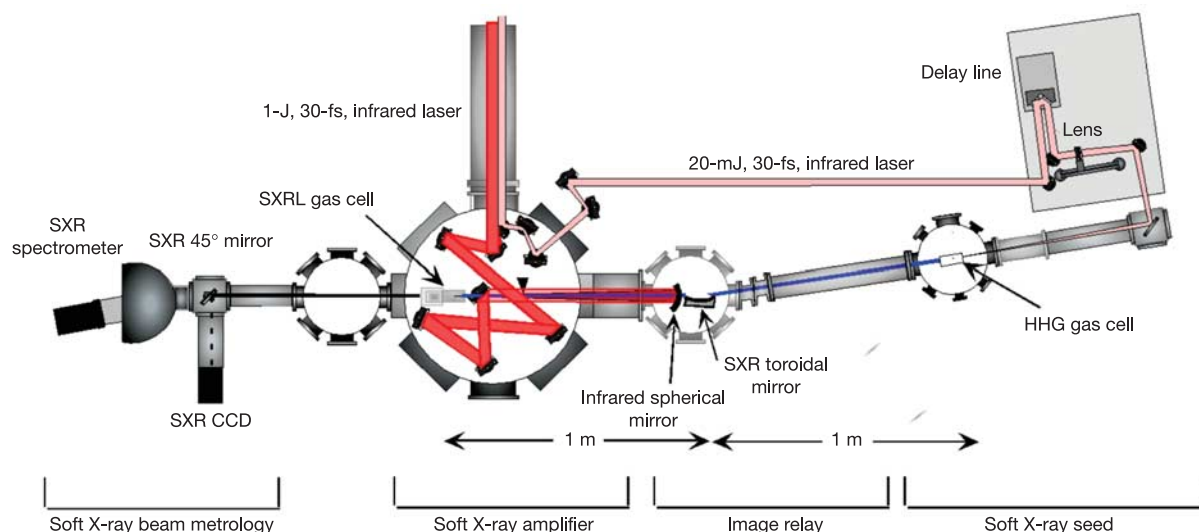


Figure 1 Schematic drawing of the experiment set-up, showing the three parts (seed, image relay and amplifier) of the soft X-ray amplifier chain.

negligible level (in terms of energy) compared to the amplified seed. Previous experiments^{2,3} aiming to test the seed/amplifier architecture for soft X-ray amplification were unable to fulfil the last two requirements. The main defects were the choice of the soft X-ray amplifier^{4,5}, a strongly refractive laser-produced plasma^{2,6} from a solid target resulting in a serious beam deterioration, and inefficient coupling between the seed and the amplifier which led to production of an amplified seed beam with a lower energy than the self-emission.

Our approach overcomes these problems by, first, using a high harmonic generation (HHG) seed⁷, second, focusing it at high intensity to increase the level of injection in order to reach the energy extraction regime, and finally, using an optical field ionized (OFI) plasma as the soft X-ray amplifying medium, of low density and thus little refraction⁸. The experimental set-up is shown on Fig. 1. The seed beam was obtained by focusing a 20-mJ, 30-fs, infrared laser in a gas cell filled with argon⁷. This seed was image-relayed onto the entrance of the soft X-ray laser (SXRL) amplifier (Kr gas longitudinally pumped by a 1-J, 30-fs laser) by means of a toroidal soft X-ray mirror. This SXRL operates on the $3d^9 4d_{J=0} \rightarrow 3d^9 4p_{J=1}$ transition of the Kr^{8+} ion at 32.8 nm (ref. 7), which has a close match with the 25th harmonic of the infrared laser. After amplification, the output beam was either analysed by a soft X-ray spectrometer or a monochromatic cross-section was recorded (see Methods section).

The delay between SXRL plasma creation and HHG injection was varied to reach optimum amplification conditions, synchronizing HHG seeding with maximum gain. Recorded emission spectra are displayed in Fig. 2, corresponding to: high harmonics alone (spectrum a), SXRL-amplified spontaneous emission (ASE) (spectrum b), SXRL with HHG seeding long after gain extinction (spectrum c), and HHG synchronized with maximum SXRL gain (spectrum d). The amplification period starts 5 ps after the interaction of the infrared laser with the gas and lasts about 8 ps. Figure 2d clearly shows that strong amplification of the seed has been achieved. In this case, HHG were seeded at about twice the intensity of the ASE of the SXRL, leading to an enhancement of the output signal by a factor of 13.

The amplification law (output intensity versus amplifier length) of the amplified spontaneous emission of the SXRL alone was measured to clarify the relationship between seeding level and saturation intensity (see Supplementary Information). The gas cell length ranged from 0 to 4 mm and the intensities were fitted

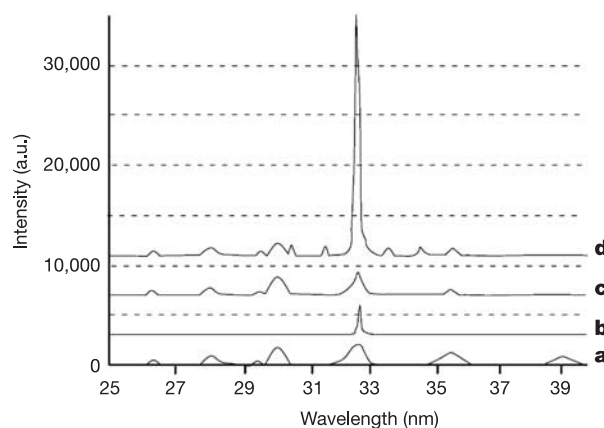


Figure 2 Experimental spectra under various conditions. **a**, HHG only; **b**, ASE SXRL only; **c**, both HHG and SXRL in a non-amplifying timing; and **d**, the amplified seeded SXRL. Each spectrum corresponds to signal accumulated over ten laser shots, to reduce the shot-to-shot error. The spectra have been artificially shifted from the bottom to the top to aid comparison. On spectrum **d**, small spikes are visible mainly around the seeded SXRL peak, as coming from its diffraction on the grating holder.

taking into account saturation effects⁹. A small-signal gain coefficient of 80 cm^{-1} has been inferred, and saturation was reached beyond 1.7 mm of SXRL plasma. In Fig. 2d, HHG seeding was well above saturation (about 4.5 times higher), that is, the 4-mm-long SXRL plasma was working in the energy extraction regime. In the second part of the experiment, the seeding intensity was reduced down to 1/50th of the saturation intensity. The output signal was then 38 times higher than saturation intensity and the amplification factor became as large as 200. Surprisingly, for both high and low seed intensities we measured an intensity output 10 to 20 times higher than the best ASE signal, whereas from the classical theory of laser amplification the seeded SXRL intensity should be only twice the ASE intensity¹⁰. This result will be explained below while discussing the estimated pulse duration.

Because HHG has a much shorter pulse duration (20 fs) than the soft X-ray laser gain duration (~ 8 ps), the final output pulse duration is limited by the amplified spectral line shape and the

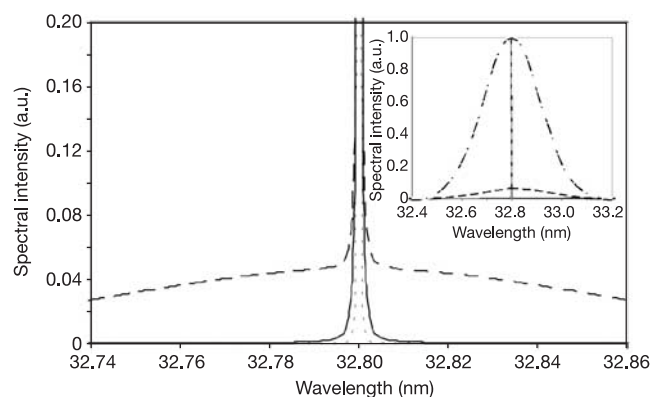


Figure 3 Calculated line shapes for the HHG (seed), the ASE SXRL, and the seeded SXRL for two levels of seeding. The inset shows the line shapes on an expanded scale pointing out the intense line wings of the seeded SXRL (dashed) as compared to the ASE SXRL (grey) and HHG (dot-dashed). The main figure is a zoom of the inset comparing the seeded SXRL line wings for the highest (dashed line) and the lowest (dotted line) seeding intensities to the intrinsic SXRL (solid line).

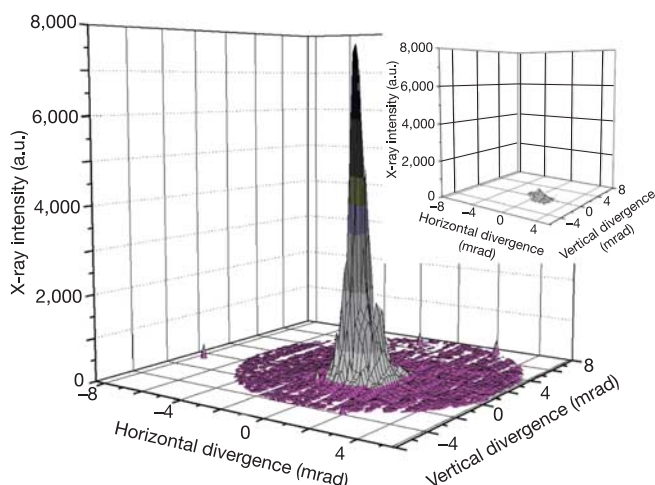


Figure 4 Three-dimensional images (false colour) of the seeded SXRL cross-section as recorded by the soft X-ray CCD detector after reflection on the monochromatic soft X-ray mirror. The central strong signal (grey) belongs to the seeded SXRL while the ASE SXRL, illuminating almost the whole filter, generates a circular signal (purple). The inset shows the seeding HHG signal recorded with the same set-up and displayed on the same scale as the main image.

spectral phase within the Fourier limit. Our previous collisional/radiative calculations¹¹ have shown that the plasma is evolving on a timescale of a few picoseconds, and so the spectral phase can be assumed to be a constant, meaning that the pulse duration depends solely on the spectral line shape. This is not the case for ASE soft X-ray lasers, where the duration of the spontaneous emission plays a major role in pulse duration¹². Two phenomena may modify the linewidth: gain narrowing and saturation re-broadening¹³. Using experimental parameters (seeding and saturation intensities as well as SXRL¹⁴ and HHG linewidths), we modelled the seeded SXRL line shape during the amplification process (radiative transfer equation¹³ with no self-emission) up to the end of the plasma. Figure 3 shows the intrinsic SXRL and HHG line shapes (lorentzian and gaussian), and the seeded SXRL line shapes after amplification over 4 mm of plasma. In both cases it appears that, owing to very

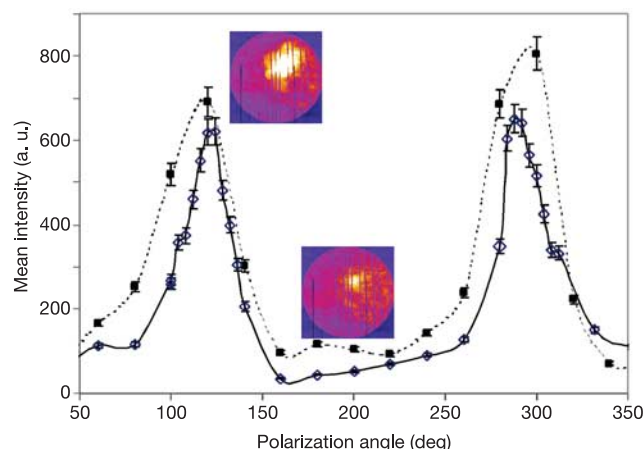


Figure 5 Variation of the HHG (open diamond) and seeded SXRL (black square) intensities versus the angle of polarization. Note that the zero angle has no physical meaning here. Insets show the seeded as well as the ASE soft X-ray laser cross-sections for s- and p-polarization. The same colour map has been used for both images. The error bars (percentage of the confidence interval) have been estimated taking into account the detector noise and the shot-to-shot energy variation of the seeded soft X-ray laser as coming from pump laser fluctuations.

strong saturation, the linewidths (at half-maximum) reach $\Delta\lambda \approx 2$ pm, implying that the Fourier-limited seeded SXRL duration is about 500 fs. To our knowledge, the seeded soft X-ray laser reported here is the only femtosecond-range soft X-ray laser demonstrated to date. Figure 3 also shows that for the lowest seeding intensity, the amplified spectrum kept the lorentzian shape because saturation is reached only at the very end of the plasma. In contrast, for the above-saturation-intensity seeding, the centre of the line is weakly amplified while the line 'wings' remain unsaturated and are strongly amplified. This is only possible because the HHG seed has a very wide linewidth, about 200 times larger than the SXRL width. Such a phenomenon could not appear for any ASE soft X-ray laser in the role of a seed. Considering the energy contained in the wings, the output signal while seeded above saturation intensity is estimated to be higher by a factor of about 15 than the ASE soft X-ray laser, in good agreement with the experimental data. As has been demonstrated both theoretically and experimentally, seeding a soft X-ray laser with a source having a much wider bandwidth is very beneficial for energy extraction. Ions with stimulated transitions that are normally spectrally shifted too far to amplify incident radiation may now participate in the amplification.

The seeded SXRL beam was also characterized in terms of optical parameters such as divergence, spatial coherence and polarization. These parameters were directly inferred from images of the beam cross-section (Figs 4 and 5). Theoretically, the ratio of the mirror reflectivity for the 25th harmonic (32.8 nm) to the 23rd (35.65 nm) or to the 27th (30.37 nm) is about ten. Therefore, only the 25th harmonic and/or the SXRL might be reflected towards the detector. The seed divergence is imposed by both the geometry of the toroidal mirror and the HHG intrinsic divergence. The divergence of the seeded SXRL was slightly lower than that of the seed. The brightest part of the image in Fig. 4 corresponds to the seeded SXRL whereas the weak, surrounding, circular signal is due to the ASE soft X-ray laser. The divergence of the ASE soft X-ray laser was found to be much larger (>12 mrad) than that of the seeded beam (1 mrad).

The multilayer mirror has been designed to be highly polarizing. The extinction ratio (intensity peak/lowest intensity) has been investigated by measuring the reflected signal versus the angle of

polarization of the HHG. Figure 5 displays the variation of both the HHG and the seeded SXRL intensity versus the angle of polarization. The extinction ratio of the mirror is as high as 20. From these curves, it appears that the seeded SXRL is polarized and follows the initial polarization of HHG well. No sign of depolarization has been observed. The insets in Fig. 5 show the beam cross-section when the HHGs were s-polarized (peak reflectivity) and p-polarized (minimum reflectivity). As expected, the ASE SXRL alone is unpolarized.

By comparing the images in Fig. 5, one may estimate the ratio between the seeded SXRL and the ASE total energies. When the seed was heavily saturated, the energy contained in the seeded SXRL was about 17 times higher than the energy of the ASE SXRL. Note that the ASE level might easily be reduced to a negligible level by using an aperture that selects only the seeded beam. Knowing the ASE level, we estimated the seeded SXRL photon number to be about 1.5×10^{11} (0.7 μ J) per pulse, and about 35 nJ for the seed. On all the images of seeded SXRL cross-section, strong diffraction structures appeared close to alignment wires. For HHG and seeded SXRL the diffraction was highly contrasted, showing that the seeded SXRL kept the high degree of coherence of the HHG. The ASE soft X-ray laser alone generated hardly visible diffraction fringes, confirming the expected weak coherence. From previous measurements of the ASE SXRL wavefront¹⁵ we may expect to maintain the seeded SXRL as well as the HHG wavefront¹⁶. Then, assuming a pulse duration of 500 fs and a focal spot close to the diffraction limit (0.1 μ m), such a seeded soft X-ray laser may achieve an intensity as high as 1.5×10^{16} W cm⁻².

Future work will aim to increase the energy of the seeded SXRL up to 0.1–1 mJ and to reduce its duration down to 100 fs. To do this requires workers both to lengthen the amplifying medium and increase its density. Capillaries filled with gas as the amplifying medium seem very promising. Indeed, recent experiments¹⁷ have shown a dramatic enhancement (30 times) of the SXRL output energy thanks to the use of capillaries instead of gas cells. By improving the geometrical coupling between the seed and the amplification zone, an extra enhancement factor of nine might be achieved on the output energy, leading to about 0.18 mJ as a first step of improvement. We note that seeding the amplifier at higher intensity than during the reported experiment should further reduce the contribution of the ASE beam to the total energy. Within a few years, we may realistically expect to produce intensities around 10^{20} W cm⁻² at around 10 nm. The seeding technique is likely to be extended down to the 'water window' (2.2–4.4 nm) because both HHG¹⁸ and SXRL¹⁹ have been demonstrated in this spectral range. As this technique matures, future compact soft X-ray laser chains should achieve beam performances close to those of soft X-ray free-electron lasers¹ and pave the way towards emerging biological applications. □

Methods

The experiment has been performed at Laboratoire d'Optique Appliquée with the multiterawatt, 30-fs, 10-Hz, Ti:sapphire laser emitting at 815 nm. The laser delivers two chirped beams having separate pulse compressor enabling one to optimize the SXRL or the HHG drivers independently (Fig. 1). One beam, often called the 'pump' beam, contains most of the energy (about 1 J on target) and is used for creating the amplifying SXRL medium. The second beam, called the 'probe' beam, has an energy of around 20 mJ on target and generates the HHG.

Harmonics were produced by focusing the probe beam into a gas cell of variable pressure and length, filled with argon. The linearly polarized beam was focused with a focal length $f = 1.5$ m, spherical lens. The focal spot has been measured to be about 150 μ m, leading to an intensity of about 10^{15} W cm⁻². The harmonic flux was optimized by changing the conditions of laser interaction with the active medium, such as moving the focal spot position, changing the gas pressure and the cell length. The highest HHG flux was obtained for a pressure of 25 mbar, a 10-mm-long cell, and a focal plane situated 4 cm behind the gas cell. The main optimization was the fine adjustment of the wavelength of one harmonic (25th) so as to coincide with the wavelength of Kr⁸⁺ SXRL (32.8 nm). This was done on a day-to-day basis by adjusting the chirp of the laser generating the HHG^{20,21}.

The amplifying plasma was created by focusing the pump beam into another gas cell of

variable length (from 0 to 6 mm) and pressure. The cell was filled with krypton held at uniform pressure. Fundamental radiation from the Ti:sapphire laser was circularly polarized by a quarter-wave plate and focused with an $f = 0.85$ m, spherical, on-axis mirror into the gas cell. Circular polarization ensures the creation of energetic free electrons that collisionally pump the Kr⁸⁺ lasing ions⁵. The measured focal spot was about 50 μ m in diameter, leading to a net intensity of about 5×10^{17} W cm⁻².

The output of the harmonic source was re-imaged on the entrance of the amplifier plasma, using a 1:1 grazing incidence (4°) gold-coated toroidal mirror. The theoretical reflectivity is around 80%. The focal spot of HHG has been measured to be around 150 μ m leading to a spatial coupling of about 1/9 between HHG and the SXRL gain region. Special attention has been paid to the temporal and spatial overlapping between the harmonic and the soft X-ray laser beams. Because the OFI SXRL is longitudinally pumped, the position of the gain region coincides with the position of the infrared driving laser. The HHG are also collinear with the infrared beam, so the superposition of the HHG and the SXRL gain region has been done in infrared. Temporal synchronization of the HHG with the period of SXRL amplification could not be done as simply because the plasma starts to amplify only a few picoseconds¹² after the interaction of the driving laser with the gas. Because HHG are emitted only during the interaction of the infrared laser with the gas, the synchronization of the seed and the SXRL period of amplification was done by adjusting the delay between the infrared pump and probe beams. A delay line with an accuracy of 30 fs was used for the synchronization of the seed with the period of amplification.

The soft X-ray beams generated in this experiment were analysed by an on-axis soft X-ray spectrometer⁹ and by making monochromatic images of the beam cross-section. These images were obtained by placing a removable 45°, soft X-ray mirror in the beam path, redirecting the beam towards a cooled, thin, back-illuminated charge-coupled device (CCD). A 300-nm-thick aluminium filter has been used to remove the infrared beams. The interferential multilayer mirror has been designed and manufactured at the Institut d'Optique Théorique et Appliquée (France) so as to achieve a high reflectivity (estimated at 50%) and a strong polarization at 45°. The soft X-ray mirror is made of a stack of 30 Mo/B₄C/Si tri-layers.

Received 29 April; accepted 19 July 2004; doi:10.1038/nature02883.

1. Ayvazyan, A. *et al.* A new powerful source for coherent VUV radiation: Demonstration of exponential growth and saturation at the TTF free-electron laser. *Eur. Phys. J. D* **20**, 149–156 (2002).
2. Ditmire, T. *et al.* Amplification of xuv harmonic radiation in a gallium amplifier. *Phys. Rev. A* **51** (6), R4337–R4340 (1995).
3. Rus, B. *et al.* Demonstration of amplification of a polarized soft-x-ray laser beam in a neonlike germanium plasma. *Phys. Rev. A* **51** (3), 2316–2327 (1995).
4. Rocca, J. J. Table top soft x-ray lasers. *Rev. Sci. Instrum.* **70**, 3799–3827 (1999).
5. Daido, H. Review of soft x-ray laser researches and developments. *Rep. Prog. Phys.* **65**, 1513–1575 (2002).
6. Plowes, J. A., Pert, G. J. & Holden, P. B. Refraction compensation by target curvature in X-ray laser. *Opt. Commun.* **116**, 260–268 (1995).
7. Kazamias, S. *et al.* Global optimization of high harmonic generation. *Phys. Rev. Lett.* **90**, 193901 (2003).
8. Sebban, S. *et al.* Demonstration of a Ni-like Kr optical-field-ionization at 32.8 nm. *Phys. Rev. Lett.* **89**, 253901 (2002).
9. Sebban, S. *et al.* Saturated amplification of a collisionally pumped optical-field-ionization soft X-ray laser at 41.8 nm. *Phys. Rev. Lett.* **86**, 3004–3007 (2001).
10. Siegman, A. E. *Lasers* (Univ. Science Books, Mill Valley, California, 1986).
11. Sebban, S. *et al.* Investigations of collisionally pumped optical field ionization soft-x-ray lasers. *J. Opt. Soc. Am. B* **20** (1), 195–202 (2003).
12. Smith, R. *et al.* Longitudinal coherence measurements of the transient collisional X-ray laser. *Opt. Lett.* **28** (22), 2261–2263 (2003).
13. Casperson, L. W. & Yariv, A. Spectral narrowing in high-gain lasers. *IEEE J. Quant. Electr.* **8** (2), 80–85 (1972).
14. Lemoff, B. E., Barty, C. P. J. & Harris, S. E. Femtosecond-pulse-driven, electron-excited XUV lasers in eight-times-ionized noble gases. *Opt. Lett.* **19** (8), 569–571 (1994).
15. Mocek, T. *et al.* Characterization of collisionally pumped optical-field-ionization soft x-ray lasers. *Appl. Phys. B* **78**, 939–944 (2004).
16. Takahashi, E. *et al.* Generation of high-energy high-order harmonics by use of a long interaction medium. *J. Opt. Soc. Am. B* **20** (1), 158–165 (2003).
17. Mocek, T. *et al.* Significant improvement of the 41.8 nm Xe⁸⁺ laser using gas-filled capillary tubes. In *Proc. 9th Int. Conf. on X-ray Lasers, 24–28 May 2004, Beijing, China* (in the press).
18. Gibson, E. *et al.* High-order harmonic generation up to 250 eV from highly ionized Argon. *Phys. Rev. Lett.* **92** (3), 033001 (2004).
19. McGowan, B. J. *et al.* Demonstration of x-ray amplifiers near the carbon K edge. *Phys. Rev. Lett.* **65**, 420–423 (1990).
20. Lee, D. G., Kim, J. H., Hong, K. H. & Nam, C. H. Coherent control of high-order harmonics with chirped femtosecond laser pulses. *Phys. Rev. Lett.* **87** (24), 243902 (2001).
21. Reitze, D. H. Enhancement of high-order harmonic generation at tuned wavelengths through adaptive control. *Opt. Lett.* **29** (1), 86–88 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. L. Charles for technical support. T.M. was supported by the Marie Curie Individual Fellowship of the European Commission.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.Z. (philippe.zeitoun@ensta.fr).

Demonstration of a quantum teleportation network for continuous variables

Hidehiro Yonezawa^{1,2}, Takao Aoki^{1,2} & Akira Furusawa^{1,2}

¹Department of Applied Physics, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

²CREST, Japan Science and Technology Agency, 1-1-9 Yaesu, Chuo-ku, Tokyo 103-0028, Japan

Quantum teleportation^{1–8} involves the transportation of an unknown quantum state from one location to another, without physical transfer of the information carrier. Although quantum teleportation is a naturally bipartite process, it can be extended to a multipartite protocol known as a quantum teleportation network⁹. In such a network, entanglement is shared between three or more parties. For the case of three parties (a tripartite network), teleportation of a quantum state can occur between any pair, but only with the assistance of the third party. Multipartite quantum protocols are expected to form fundamental components for larger-scale quantum communication and computation^{10,11}. Here we report the experimental realization of a tripartite quantum teleportation network for quantum states of continuous variables (electromagnetic field modes). We demonstrate teleportation of a coherent state between three different pairs in the network, unambiguously demonstrating its tripartite character.

A quantum teleportation network is a quantum communication network linked by quantum teleportation. For example, in a tripartite network, three parties (which we call Alice, Bob and Claire) are connected on the network in which they are spatially separated and over which they have previously shared tripartite entanglement. They can only use local operations and classical channels to communicate with each other. In principle, they need not even know where the others are, as long as they can communicate through the classical channels. Under these conditions, they exchange a quantum state. Here the quantum state to be teleported is that of an electromagnetic field mode. We use the Heisenberg picture to describe the evolution of the quantum state. An electromagnetic field mode is represented by an annihilation operator $\hat{a}$ whose real and imaginary parts ($\hat{a} = \hat{x} + i\hat{p}$) correspond to the position and momentum quadrature-phase amplitude operators. These operators $\hat{x}$ and $\hat{p}$ satisfy the commutation relation $[\hat{x}, \hat{p}] = i/2$ (units-free, with $\hbar = 1/2$). In some respects, a quantum teleportation network is similar to bipartite quantum teleportation. In both schemes the parties share quantum entanglement, and send a quantum state using local operations and the classical channels. But the properties of tripartite entanglement make it different from bipartite teleportation in other respects.

If Alice sends a quantum state to Bob, what role does Claire play? We recall that the three parties are in the tripartite entangled state. The third party Claire also has a quantum correlation with the other parties. Thus Alice and Bob need Claire's information to succeed in teleportation. In other words, Claire can control the transfer of the quantum state from Alice to Bob by restricting their access to her information. This is a clear manifestation of tripartite entanglement.

Tripartite entanglement for continuous variables (CVs) can be generated by using three squeezed vacuum states and two beam splitters^{9,12}. In the limit of infinite squeezing, the state is the CV analogue^{9,12} of a Greenberger–Horne–Zeilinger (GHZ) state¹³. The CV GHZ state is a maximally entangled state and a simultaneous eigenstate of zero total momentum ($p_1 + p_2 + p_3 = 0$) and zero

relative positions ($x_i - x_j = 0$, $i, j = 1, 2, 3$). The entanglement properties of the GHZ state are easily destroyed if part of the system is disturbed. For example, if one of the three subsystems is traced out (discarded), the remaining state ($\hat{\rho}_{AB}, \hat{\rho}_{AC}, \hat{\rho}_{BC}$) is completely unentangled¹⁴. Thus without Claire's information the quantum entanglement between Alice and Bob vanishes, and quantum teleportation is no longer possible.

In a real experiment, a maximally entangled state is not available because of finite squeezing and inevitable losses. But we can still obtain a fully inseparable tripartite entangled state (a state none of whose subsystems can be separated)¹². An entangled state generated by three highly squeezed vacuum states still behaves like the GHZ state. The properties of the state are easily destroyed if part of the system is disturbed. In this case, Claire can completely determine success or failure of quantum teleportation between Alice and Bob. In contrast, even if three weakly squeezed vacuum states are used, the state is a fully inseparable tripartite entangled state, but the remaining bipartite state after tracing out one of the three subsystems is still entangled¹⁵. In this case, after tracing out one subsystem (for example, mode 3), the variance $\langle [\Delta(\hat{x}_1 - \hat{x}_2)]^2 \rangle + \langle [\Delta(\hat{p}_1 + \hat{p}_2)]^2 \rangle$ is still below unity¹⁵ and shows the presence of bipartite entanglement between modes 1 and 2. If we use such a state, we will succeed in teleportation even without Claire's information, although teleportation fidelity is lower than the case with her information. In order to control success or failure of teleportation, we need to use three highly squeezed vacuum states.

There is another important point to note when we develop bipartite quantum teleportation into a tripartite quantum teleportation network. Only if we use a fully inseparable tripartite entangled state can we succeed in teleportation between an arbitrary pair in the network. Namely, each party can play any of three roles: a sender, a receiver and a controller. Note that if we use a partially entangled state, we may succeed in teleportation for a particular combination of the sender, the receiver and the controller, but we may fail for other combinations. From this point of view, a truly tripartite quantum protocol is defined as a protocol that succeeds only if a fully inseparable tripartite entanglement is used. In order to verify success of a truly tripartite quantum protocol, we need to succeed in teleportation for at least two different combinations^{12,15}.

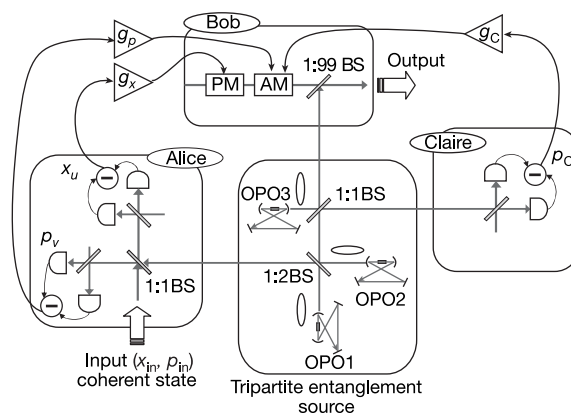


Figure 1 The experimental set-up for quantum teleportation from Alice to Bob under the control of Claire. In order to generate quadrature squeezed vacuum states, we use subthreshold optical parametric oscillators (OPOs) with a potassium niobate crystal. The output of a continuous wave Ti:sapphire laser at 860 nm is frequency doubled in an external cavity. The output beam at 430 nm is divided into three beams to pump three OPOs. The pump powers are about 50 mW. Three squeezed vacuum states are combined at two beam splitters (BS) with relative phases locked. The three outputs of the two beam splitters are in a tripartite entangled state, and are sent to the three parties Alice, Bob and Claire. The input state is a coherent state and an optical sideband at 1 MHz. Symbols and abbreviations are defined in the text.

For example, in the experiment by Jing *et al.*¹⁶, a controlled dense coding for a particular combination shows only partial success and is not sufficient for the demonstration of a truly tripartite quantum protocol. In our experiment, we demonstrate quantum teleportation for three different combinations. To the best of our knowledge, this is the first demonstration of a truly tripartite quantum protocol.

Here we describe the procedure of our quantum teleportation network experiment. Figure 1 shows a diagram of our experimental set-up. Tripartite entangled states¹² are distributed to Alice, Bob and Claire. We represent the operators for each mode as $(\hat{x}_i, \hat{p}_i)$ ($i = A, B, C$) in the Heisenberg representation. We first consider teleportation with the combination of sender Alice, receiver Bob, and controller Claire.

First, Alice performs a joint measurement or so-called ‘Bell measurement’ on her entangled mode $(\hat{x}_A, \hat{p}_A)$ and an unknown input mode $(\hat{x}_{in}, \hat{p}_{in})$. In our experiment, the input state is a coherent state and an optical sideband of continuous wave 860-nm carrier light. The Bell measurement instrument consists of a half beam splitter and two optical homodyne detectors. Two outputs of the half beam splitter are labelled as $\hat{x}_u = (\hat{x}_{in} - \hat{x}_A)/\sqrt{2}$ and $\hat{p}_v = (\hat{p}_{in} + \hat{p}_A)/\sqrt{2}$ for relevant quadratures. The outputs of the homodyne detectors are measured with a spectrum analyser. Figure 2a shows the measurement result of $\langle(\hat{p}_v)^2\rangle$ at Alice ($\langle(\hat{x}_u)^2\rangle$ is not shown). The visibility between the input modes and the local oscillators of Alice’s homodyne detectors are both 0.99. The visibilities between the other modes are about 0.97. The quantum efficiency of the photodiode is 0.998 at 860 nm. In Fig. 2a, the noise level of a vacuum input $\langle(\Delta\hat{p}_v)^2\rangle$ (the variance of $\hat{p}_v$) is 3.7 dB compared to the corresponding vacuum noise level $\langle(\Delta\hat{p}^{(0)})^2\rangle$, while the noise level for x quadrature is 2.1 dB (not shown). The measured noise levels for x and p quadratures are asymmetric. This is because tripartite entanglement is generated by two x and one p quadrature squeezed vacuum states. For example, the variances of mode A are described as $\langle(\Delta\hat{x}_A)^2\rangle = (2e^{-2r} + e^{2r})/12$, $\langle(\Delta\hat{p}_A)^2\rangle = (e^{-2r} + 2e^{2r})/12$, respectively¹² (here r is a squeezing parameter which is the same for the three squeezed states; $r = r_1 = r_2 = r_3$). Thus the measured noise levels at Alice ($\langle(\Delta\hat{p}_v)^2\rangle$ and $\langle(\Delta\hat{x}_u)^2\rangle$) are asymmetric.

The measured values x_u and p_v for $\hat{x}_u$ and $\hat{p}_v$ are sent through the classical channels with gain g_x and g_p , respectively. The third party Claire measures $\hat{p}_C$ of her entangled mode. Note that Claire does not measure the x quadrature. Figure 2b shows her measurement result

$\langle(\Delta\hat{p}_C)^2\rangle$. Claire also observes a noise and sends it to Bob through the classical channel with gain g_C .

The gains of the classical channels are adjusted in the manner of ref. 8. The normalized gains of Alice’s classical channels $g_x = \langle\hat{x}_{out}\rangle/\langle\hat{x}_{in}\rangle$, $g_p = \langle\hat{p}_{out}\rangle/\langle\hat{p}_{in}\rangle$ are adjusted to $g_x = 0.99 \pm 0.04$ and $g_p = 1.00 \pm 0.03$, respectively. For simplicity, these gains are fixed throughout the experiment and treated as unity.

Let us write Bob’s initial mode before the measurement of Alice and Claire as:

$$\begin{aligned}\hat{x}_B &= \hat{x}_{in} - (\hat{x}_A - \hat{x}_B) - \sqrt{2}\hat{x}_u \\ \hat{p}_B &= \hat{p}_{in} + (\hat{p}_A + \hat{p}_B + g_C\hat{p}_C) - \sqrt{2}\hat{p}_v - g_C\hat{p}_C\end{aligned}\quad (1)$$

Note that in this step Bob’s mode remains unchanged. After measuring $\hat{x}_u, \hat{p}_v$ and $\hat{p}_C$, these operators collapse and reduce to certain values. Receiving these measurement results, Bob displaces his mode as $\hat{x}_B \rightarrow \hat{x}_{tel} = \hat{x}_B + \sqrt{2}x_u, \hat{p}_B \rightarrow \hat{p}_{tel} = \hat{p}_B + \sqrt{2}p_v + g_C p_C$ and accomplishes the teleportation. In our experiment, displacement is performed by applying electro-optical modulations. Bob modulates a beam by using amplitude and phase modulators (AM and PM in Fig. 1). The amplitude and phase modulations correspond to the displacement of p and x quadratures, respectively. The modulated beam is combined with Bob’s mode $(\hat{x}_B, \hat{p}_B)$ at a 1/99 beam splitter.

The teleported mode becomes:

$$\begin{aligned}\hat{x}_{tel} &= \hat{x}_{in} - (\hat{x}_A - \hat{x}_B) \\ \hat{p}_{tel} &= \hat{p}_{in} + (\hat{p}_A + \hat{p}_B + g_C\hat{p}_C)\end{aligned}\quad (2)$$

In the ideal case, total momentum $\hat{p}_A + \hat{p}_B + \hat{p}_C$ and relative position $\hat{x}_A - \hat{x}_B$ have zero-eigenvalues $p_A + p_B + p_C = 0$ and $x_A - x_B = 0$ simultaneously, and the teleported state is identical to the input state ($g_C = 1$). In a real experiment, however, the teleported state has additional fluctuations. Without entanglement, at least two units of a vacuum fluctuation are added ($g_C = 0$). These additional vacuum fluctuations are called two qudities¹⁷ which must be paid for crossing the boundary between classical and quantum domains.

Figure 2c shows measurement results of the teleported mode for p quadrature with gain $g_C = 1.02 \pm 0.03$. The noise level of a vacuum input $\langle(\Delta\hat{p}_{tel})^2\rangle$ is 3.3 dB compared to the corresponding vacuum noise level for p quadrature, while the noise level for x quadrature is

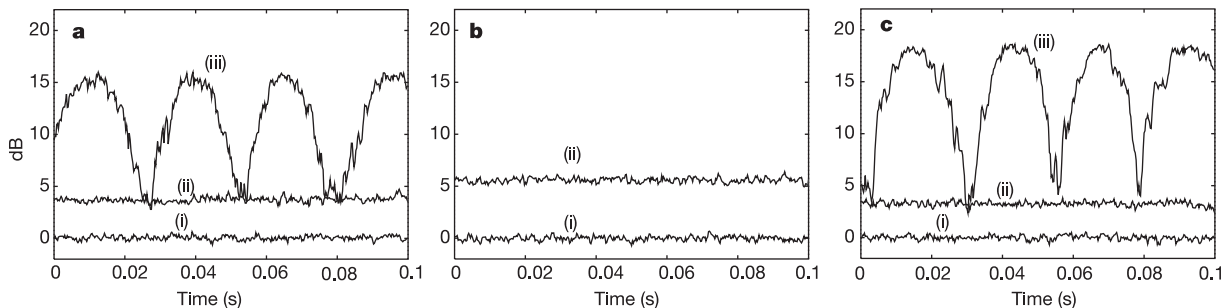


Figure 2 Quantum teleportation from Alice to Bob under the control of Claire. All traces are normalized to the corresponding vacuum noise levels. **a**, Alice’s measurement results for p quadrature (x quadrature is not shown). Trace i, the corresponding vacuum noise level $\langle(\Delta\hat{p}^{(0)})^2\rangle = 1/4$. Trace ii, the measurement result of a vacuum input $\langle(\hat{p}_v')^2\rangle$ where $\hat{p}_v' = (\hat{p}_{in}^{(0)} + \hat{p}_A)/\sqrt{2}$. Note that $\langle\hat{p}_{in}^{(0)}\rangle = \langle\hat{p}_A\rangle = 0$; thus $\langle(\hat{p}_v')^2\rangle = \langle(\Delta\hat{p}_v')^2\rangle$, and the variance of a coherent state input $\langle(\Delta\hat{p}_v)^2\rangle$ is equivalent to $\langle(\Delta\hat{p}_v')^2\rangle$ because a vacuum state is a coherent state with zero amplitude. The noise level is 3.7 dB, while the noise level for x quadrature is 2.1 dB (not shown). Trace iii, the measurement result of a coherent state input $\langle(\hat{p}_v)^2\rangle$ with the phase scanned. The measured amplitude of the input beam is about 15 dB, which is 3 dB lower than the actual input because of the input half beam splitter. **b**, Claire’s measurement results for p quadrature. Trace i, the

corresponding vacuum noise level. Trace ii, the measurement result of $\langle(\hat{p}_C)^2\rangle = \langle(\Delta\hat{p}_C)^2\rangle$. The noise level is 5.7 dB. **c**, The measurement results of the teleported states for p quadrature (x quadrature is not shown). Trace i, the corresponding vacuum noise level. Trace ii, the teleported state for a vacuum input $\langle(\Delta\hat{p}_{tel})^2\rangle$. Note that the variance of the teleported state for a vacuum input corresponds to that for a coherent state input. The noise level is 3.3 dB for p quadrature, while the noise level for x quadrature is 3.5 dB (not shown). Trace iii, the teleported state for a coherent state input. The measured amplitude of the teleported state is about 18 dB, which is 3 dB higher than that of the input at Alice. It shows that the classical channel’s gains are almost unity. The measurement frequency is centred at 1 MHz, and the resolution and video bandwidths are 30 kHz and 300 Hz, respectively. All traces except for trace iii are averaged ten times.

3.5 dB (not shown). In the classical limit, the teleported state has three units of a vacuum fluctuation (one unit for the input state, the other two for quduties), and the variances of the teleported state become 4.8 dB compared to the corresponding vacuum noise level. The observed noise reduction from the classical limit shows the success of teleportation.

To evaluate the performance of teleportation, we use a fidelity which is defined as $F = \langle \psi_{\text{in}} | \rho_{\text{out}} | \psi_{\text{in}} \rangle$ (refs 18, 19). Although the classical limit of teleportation is derived for the case of two parties in ref. 18, it can be applied to the case of three parties⁹. In a classical case, the three parties have no quantum correlation with each other. Thus the third party cannot improve the performance of teleportation beyond the classical limit. In ref. 18, the classical limit of teleportation is derived by averaging the fidelity for a randomly chosen coherent state input. The classical limit of the averaged fidelity F_{av} is 1/2. In a real experiment, however, it is impossible to take an average over the entire phase space. But if the gains of the classical channels are unity $g_x = g_p = 1$ (not including g_C), the averaged fidelity is identical to the fidelity for a particular coherent state input ($F_{\text{av}} = F$) (ref. 9). This is because the fidelity with unity

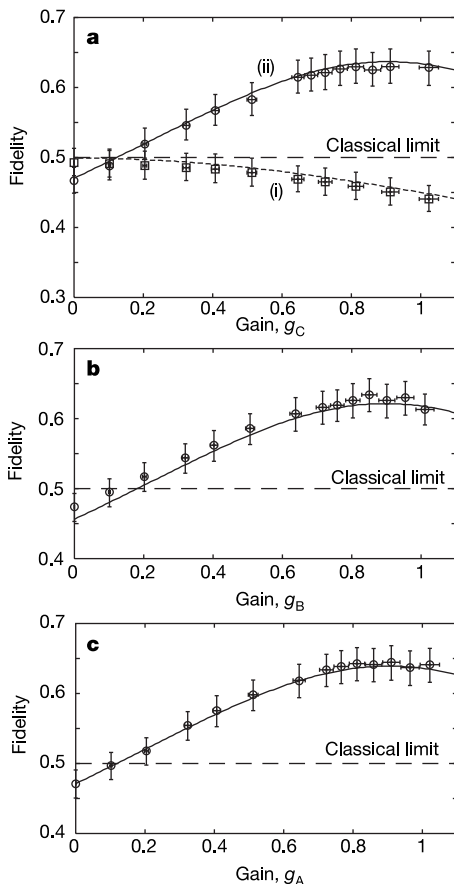


Figure 3 Dependence of the fidelities on the controller's gain. **a**, The fidelity of the teleportation from Alice to Bob under the control of Claire. Trace i, the teleportation without entanglement. Trace ii, the teleportation with tripartite entanglement. The best fidelity (0.63 ± 0.02) is obtained at $g_C \approx 0.9$ and the fidelity is better than 1/2 in the broad range of g_C . The solid lines represent the theoretical curves calculated from the experimental conditions. **b**, The fidelity of the teleportation from Alice to Claire under the control of Bob. The best fidelity (0.63 ± 0.02) is obtained at $g_B \approx 0.9$. **c**, The fidelity of the teleportation from Claire to Bob under the control of Alice. The best fidelity (0.64 ± 0.02) is obtained at $g_A \approx 0.9$. In Fig. 2 each trace contains 401 measurement points, and the measurement series are repeated three times. Error bars show 1σ , calculated using 401 measurement points and averaged over three times measurements.

gains can be determined by only the variances of the teleported state, that is, independent of the amplitude of the coherent state input. The fidelity for a coherent state input can be written $F = 2/\sqrt{1 + 4\sigma_x}\sqrt{1 + 4\sigma_p}$, where $\sigma_x = \langle (\Delta\hat{x}_{\text{tel}})^2 \rangle$ and $\sigma_p = \langle (\Delta\hat{p}_{\text{tel}})^2 \rangle$ (ref. 9). Note that a coherent state is a state of minimum uncertainty whose variances are $\langle (\Delta\hat{x}^{(0)})^2 \rangle = \langle (\Delta\hat{p}^{(0)})^2 \rangle = 1/4$. We set the gains $g_x = g_p = 1$, and examine the teleportation apparatus for a particular coherent state input.

Although we set the gains $g_x = g_p = 1$ to estimate fidelity, we change the third party's gain g_C . The best fidelity is obtained at an optimum value of g_C which is determined by the degree of squeezing⁹. We measure the fidelity, and examine the dependence of the fidelity on gain g_C .

The fidelity calculated from the variances of the teleported state is plotted as a function of g_C in Fig. 3a. Without entanglement, the fidelity is lower than 1/2. Quantum teleportation fails and optimum g_C is zero because Claire has no correlation with the other parties. With tripartite entanglement, we obtain $F = 0.63 \pm 0.02$ ($g_C \approx 0.9$) which clearly shows success of quantum teleportation between Alice and Bob. At $g_C = 0$, however, quantum teleportation fails. This is because the tripartite entanglement used in this experiment behaves like the GHZ state. To succeed in teleportation, Alice and Bob need Claire's information. If Claire does not send her information to them, the fidelity becomes even lower than that without entanglement. This clearly shows that Claire controls success or failure of the teleportation.

So far we have demonstrated the experiment for the particular combination of sender Alice, receiver Bob and controller Claire. Note again that we need to perform experiments for at least two different combinations to verify success of a truly tripartite quantum protocol. We now perform experiments of teleportation for two other combinations; one is sender Alice, receiver Claire and controller Bob and the other is sender Claire, receiver Bob and controller Alice. We change the configuration of the experimental set-up only locally while the global configuration remains unchanged. Specifically, the paths distributing the tripartite entangled states remain unchanged throughout the experiment. On the other hand, each party changes his or her set-up locally according to their roles.

The gain dependence of the teleportation fidelity from Alice to Claire and from Claire to Bob are shown in Fig. 3b and c, respectively. Both figures show almost the same dependence as Fig. 3a. This shows that our tripartite entanglement source has the same capability to perform teleportation for different combinations. The best fidelities are 0.63 ± 0.02 and 0.64 ± 0.02 ($g_B \approx g_A \approx 0.9$), respectively, which are greater than the classical limit $F = 1/2$ and show success of teleportation for these two combinations. In total, we have demonstrated three different combinations. These results show the success of a quantum teleportation network, that is, a truly tripartite quantum protocol.

The techniques used in this experiment are easily extendable to other quantum protocols, such as optimal telecloning²⁰, controlled dense coding or a quantum network containing more than three parties⁹. Multipartite quantum protocols are of practical importance in realizing more complicated quantum computation and quantum communication among many parties. □

Received 14 May; accepted 19 July 2004; doi:10.1038/nature02858.

1. Bennett, C. H. *et al.* Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
2. Bouwmeester, D. *et al.* Experimental quantum teleportation. *Nature* **390**, 575–579 (1997).
3. Boschi, D., Branca, S., De Martini, F., Hardy, L. & Popescu, S. Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **80**, 1121–1125 (1998).
4. Nielsen, M. A., Knill, E. & Laflamme, R. Complete quantum teleportation using nuclear magnetic resonance. *Nature* **396**, 52–55 (1998).
5. Marcikic, I., De Riedmatten, H., Tittel, W., Zbinden, H. & Gisin, N. Long-distance teleportation of qubits at telecommunication wavelengths. *Nature* **421**, 509–513 (2003).
6. Furusawa, A. *et al.* Unconditional quantum teleportation. *Science* **282**, 706–709 (1998).

7. Bowen, W. P. *et al.* Experimental investigation of continuous-variable quantum teleportation. *Phys. Rev. A* **67**, 032302 (2003).
8. Zhang, T. C., Goh, K. W., Chou, C. W., Lodahl, P. & Kimble, H. J. Quantum teleportation of light beams. *Phys. Rev. A* **67**, 033802 (2003).
9. van Loock, P. & Braunstein, S. L. Multipartite entanglement for continuous variables: A quantum teleportation network. *Phys. Rev. Lett.* **84**, 3482–3485 (2000).
10. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, 2000).
11. Braunstein, S. L. & Pati, A. K. *Quantum Information with Continuous Variables* (Kluwer Academic, Dordrecht, 2003).
12. Aoki, T. *et al.* Experimental creation of a fully inseparable tripartite continuous-variable state. *Phys. Rev. Lett.* **91**, 080404 (2003).
13. Greenberger, D. M., Horne, M. A., Shimony, A. & Zeilinger, A. Bell's theorem without inequalities. *Am. J. Phys.* **58**, 1131–1143 (1990).
14. Dür, W., Vidal, G. & Cirac, J. I. Three qubits can be entangled in two inequivalent ways. *Phys. Rev. A* **62**, 062314 (2000).
15. van Loock, P. & Furusawa, A. Detecting genuine multipartite continuous-variable entanglement. *Phys. Rev. A* **67**, 052315 (2003).
16. Jing, J. *et al.* Experimental demonstration of tripartite entanglement and controlled dense coding for continuous variables. *Phys. Rev. Lett.* **90**, 167903 (2003).
17. Braunstein, S. L. & Kimble, H. J. Teleportation of continuous quantum variables. *Phys. Rev. Lett.* **80**, 869–872 (1998).
18. Braunstein, S. L., Fuchs, C. A. & Kimble, H. J. Criteria for continuous-variable quantum teleportation. *J. Mod. Opt.* **47**, 267–278 (2000).
19. Braunstein, S. L., Fuchs, C. A., Kimble, H. J. & Loock, P. Quantum versus classical domains for teleportation with continuous variables. *Phys. Rev. A* **64**, 022321 (2001).
20. van Loock, P. & Braunstein, S. L. Telecloning of continuous quantum variables. *Phys. Rev. Lett.* **87**, 247901 (2001).

Acknowledgements This work was partly supported by the MEXT and the MPHPT of Japan, and the Research Foundation for Opto-Science and Technology.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.F. (akiraf@ap.t.u-tokyo.ac.jp).

A Taylor vortex analogy in granular flows

Stephen L. Conway^{1*}, Troy Shinbrot² & Benjamin J. Glasser¹

¹Department of Chemical and Biochemical Engineering, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, USA

²Department of Biomedical Engineering, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, USA

* Present address: Merck & Co., Inc., Whitehouse Station, New Jersey 08889, USA

Fluids sheared between concentric rotating cylinders undergo a series of three-dimensional instabilities. Since Taylor's archetypal 1923 study¹, these have proved pivotal to understanding how fluid flows become unstable and eventually undergo transitions to chaotic or turbulent states^{2–5}. In contrast, predicting the dynamics of granular systems—from nano-sized particles to debris flows—is far less reliable. Under shear these materials resemble fluids, but solid-like responses, non-equilibrium structures and segregation patterns develop unexpectedly^{6–9}. As a result, the analysis of geophysical events¹⁰ and the performance of largely empirical particle technologies might suffer^{11,12}. Here, using gas fluidization to overcome jamming^{6,13}, we show experimentally that granular materials develop vortices consistent with the primary Taylor instability in fluids. However, the vortices observed in our fluidized granular bed are unlike those in fluids in that they are accompanied by novel mixing–segregation transitions. The vortices seem to alleviate increased strain by spawning new vortices, directly modifying the scale of kinetic interactions. Our observations provide insights into the

mechanisms of shear transmission by particles and their consequent convective mixing.

In the Taylor experiment, uniform, initially azimuthal fluid flows spontaneously make a transition to highly ordered patterns once the inner cylinder rotation rate exceeds a critical value. The primary instability occurs when the centrifugal force generated in a rotating fluid overcomes viscosity to cause the fluid to develop counter-rotating laminar tori¹, while subsequent instabilities instigate an assortment of secondary regimes, from wavy vortices to spiral turbulence^{2,14}. Detailed investigation of these instabilities has been central to a comprehensive understanding of fluid flows and associated mixing and transport rates. By contrast, despite previous attempts, analogous instabilities have not been reported in granular flows¹⁵, which are arguably as important as their fluid counterparts. Here we examine the hypothesis that the difficulty in generating these instabilities is associated with jamming and stress chains long known to characterize granular beds^{6,13}. To defeat these stress chains, we weakly fluidize granular beds by gently pumping air uniformly upwards from below¹⁶, and find that Taylor-like tori are indeed recovered, albeit with some uniquely granular characteristics.

In Fig. 1 we show patterns observed in side views of a gas-fluidized granular bed sheared in the Taylor–Couette geometry^{15,17}. As the inner cylinder rotational velocity, Ω , is slowly incremented (by no more than 15 r.p.m. min^{−1}), a binary mixture of glass spheres spontaneously segregates into ordered, nearly axisymmetric bands, each consisting of a stripe of large red particles next to a stripe of small white particles. To the right of each snapshot we display vertical profiles of concentration obtained from digital images of the bed, where particle colour differences are processed¹⁸ to identify flow patterns.

To provide both an evaluation of the influence of fluidization and a basis for comparison with fluids, we characterize the rheology of the granular bed. As depicted in Fig. 2a (left), air flows through the annular space to fluidize particles supported on a distributor plate. The upper surface is unconfined. Flow rates are maintained below the bubbling transition^{15–17}, and the bed pressure drop is small compared with that across the distributor¹⁶. We have examined various annular gaps, d , between the cylinders but find that banding occurs most readily when the ratio of inner to outer cylinder radii is 0.7, corresponding to superficial gas velocities of 40–120 mm s^{−1}. We measure both bed shear stress, τ , and an effective bed viscosity, μ_{eff} , for various gas flows in a standard manner¹⁹ (Fig. 2a, b). We find that shear stress is nearly independent of rate and virtually identical for any non-zero airflow, and is distinct from the higher stresses generated in the unfluidized case. Viscosity exhibits a power-law dependence on shear rate, $\dot{\gamma}$. Because the fluidized-bed rheology at a particular shear rate is uncorrelated with airflow, and because we have also determined that the critical shear rate at which the patterns first appear does not change measurably for airflow variations of up to 13%, we subsequently hold airflow constant at 9 l min^{−1}.

Fluid Taylor vortices are characterized by the inner-cylinder Reynolds number, $Re_i = r_i \Omega d \rho / \mu$, where transitional Re_i vary with the aspect ratio of the Couette gap; here ρ is the density, μ is the viscosity, r_i is the radius of the rotating inner cylinder, Ω is its angular speed and d is the annular gap. Primary fluid vortices appear (in cells of similar geometry to our cell) as Re_i reaches 102–120, with destabilization to wavy vortices at $Re_i = 131$ –140 (refs 2, 14). In our fluidized granular bed, by contrast, the particulate Reynolds number, $Re_{ip} = r_i \Omega d \rho / \mu_{\text{eff}}$ is calculated using the bed bulk density, ρ , and the shear-dependent $\mu_{\text{eff}}(\Omega)$ from Fig. 2b. We note that whereas Re_i evaluated for interstitial air alone is more than 1,000 and is theoretically sufficient to initiate gas-phase instabilities that could be expected to cause banding, the state in Fig. 1d is not altered as airflow is varied from 6 to 24 l min^{−1}. Upon substitution of helium for air, transitions between banded states occur at almost

identical rotation rates even though using helium increases the gas-phase kinematic viscosity by almost a factor of eight.

Using the above definition of Re_{ip} , we identify the behaviour leading to the banding patterns as follows. From rest, we initiate shear in a well-mixed fluidized binary mixture consisting of equal parts spheres with diameters of $138\text{ }\mu\text{m}$ (white) and $462\text{ }\mu\text{m}$ (red). Particle motion is at first azimuthal and (observed at the free surface) confined to a narrow zone near the moving wall. This zone widens with rising Ω until it reaches the outer wall at $Re_{ip} \approx 26$. At this point we see an abrupt transition in which particles begin to flow radially outwards at the upper surface and downwards along the outer wall to a depth of only a few millimetres. Particle segregation by size is manifest at $Re_{ip} = 36\text{--}50$, both on the upper surface as a ring of the larger particles surrounding the inner cylinder and as subsurface axial segregation (visible through the outer wall). This behaviour is similar to segregation due to elutriation, seen industrially both in fluidized beds and in the filling of silos^{9,20}, in which smaller, lighter particles graduate to higher positions in a bed under the influence of fluid drag.

The segregation pattern shown in Fig. 1a initiates with an abrupt increase in the sharpness of the interface separating the mixture constituents. Large particles leave the free surface, and secondary circulation appears to strengthen as small particles sink significantly at the outer wall and reach the interface between large and small particle regions. Movement in the opposite sense is observed in the

lower half of the bed for large particles. Subsequent configurations appear as bands split. Transitions are rapid once any of the identified critical Reynolds numbers in Fig. 1 is exceeded, generally taking less than a few seconds. After the transition leading to Fig. 1b, two additional faint bands of large particles appear almost simultaneously, equidistant from the previous interface location. This is accompanied by mixing: composition gradients are smoothed and some large particles now breach the upper surface. In Fig. 1c–f, four, five and then seven bands emerge; concentration variations thereafter diminish as bands become indistinguishable. All segregated states appear independently of initial state (homogenous to complete vertical or horizontal segregation were investigated). Patterns return to the state in Fig. 1a with slow reduction in Ω , and can be frozen in place by abruptly halting the inner cylinder's rotation. We have investigated numerous particle size distributions and find that the combination in Fig. 1 (see legend) provides the most vivid band formation. The patterns shown are stable for at least an hour once formed (over which time, measurements indicate that mean particle size is reduced by attrition, but this is limited to about 8% when shearing at 800 r.p.m. and is not significant enough to affect band formation dynamics).

We find several features in common with the primary Taylor vortex instability, but also important differences. First, the onset Re_{ip} of 96–105 is reasonably close to the onset Re_i for fluids, 102–120 (refs 2, 14). The stripe spacing, λ , averages $0.9d$ but alternates

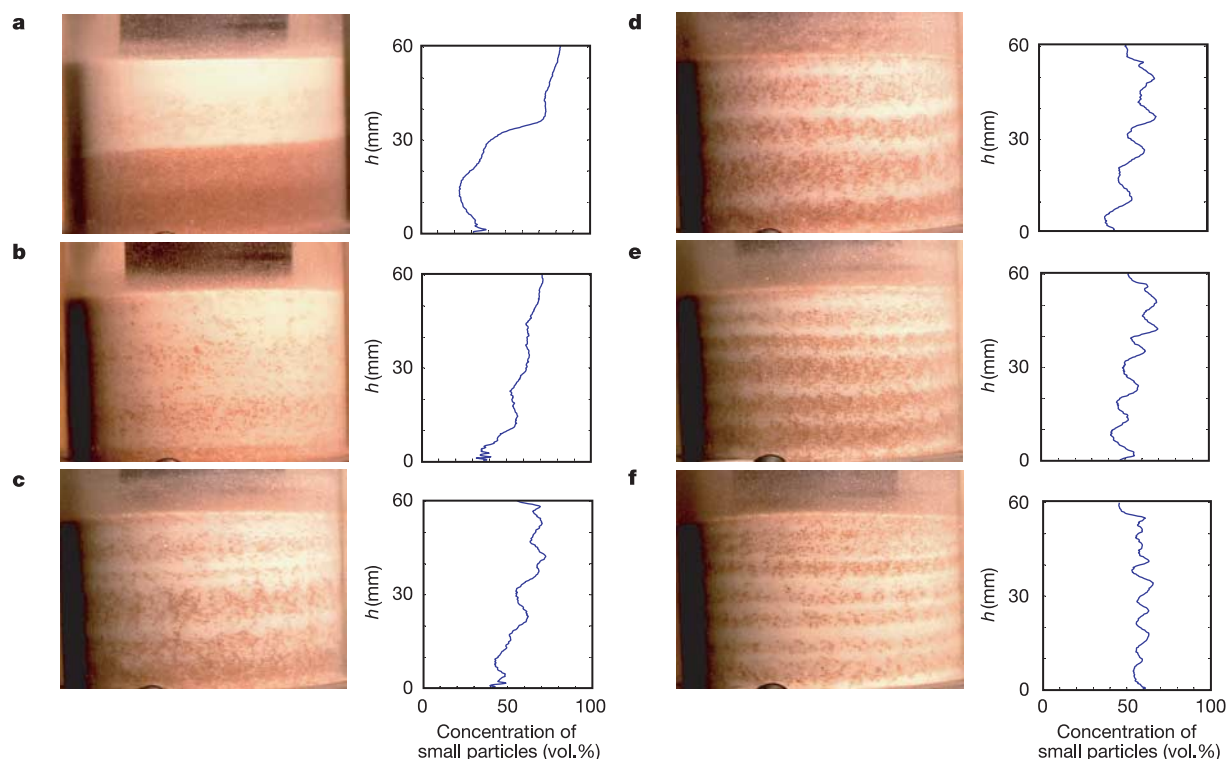


Figure 1 Fluidized granular bed imaged through a transparent outer cylinder alongside vertical composition profiles. Ω is increased from 500 r.p.m. (**a**) to 850 r.p.m. (**f**). A 50:50 (vol.%) mixture of glass spheres of diameter $140 \pm 10\text{ }\mu\text{m}$ (white) and $460 \pm 40\text{ }\mu\text{m}$ (red) is contained between concentric acrylic cylinders to a height, h , of $58 \pm 5\text{ mm}$ in an annular gap, d , of 13 mm . Fluidizing air is passed through the bed from a ceramic distributor/support at 9 l min^{-1} (standard temperature and pressure), with a maximum pressure decrease of $1,970 \pm 80\text{ Pa}$. Ambient pressure is maintained at the upper free surface. Ω is sufficient to suppress bubbling instabilities and maintain a uniform bed expansion. The inner cylinder, visible near the tops of the images, is 50 mm in diameter and coated with black 80 grit sandpaper; it rotates at $50\text{--}2,000\text{ r.p.m.}$ to generate variable shear stress. Images depict the entire bed height and width, and patterns are nearly axisymmetric. **a**, At an inner cylinder rotation of 500 r.p.m. ($Re_{ip} = 96 \pm 6$), a

single band contains large particles almost completely segregated beneath smaller particles. **b**, At 516 r.p.m. ($Re_{ip} = 105 \pm 9$), axial segregation is reduced but three faint bands now appear. **c**, At 631 r.p.m. ($Re_{ip} = 158 \pm 9$), a fourth thin band appears at $h \approx 40\text{ mm}$. **d**, At 672 r.p.m. ($Re_{ip} = 177 \pm 8$), the four bands are almost equally spaced (bands broaden with increasing Ω). **e**, At 682 r.p.m. ($Re_{ip} = 182 \pm 8$), a fifth band appears. **f**, At 850 r.p.m. ($Re_{ip} = 287 \pm 20$), seven bands are visible. For **b–f**, smaller bands also appear at the free surface and bed base. Image analysis yields corresponding vertical composition profiles, obtained by a correlation of pixel intensities by means of six-point calibration curves (repeated for both loose and tapped packings, with $R^2 = 0.9986$ achieved by using standard and background reference images under controlled lighting). Concentrations are averaged horizontally over the width of the outer cylinder.

around this value for stripes of large and small particles. This can be compared with fluid Taylor vortex spacing $\lambda \sim d$, also alternating between $0.8d$ and $1.2d$ for inflow–outflow fluid vortices²¹. Although particles have been observed to move radially at the free surface in non-fluidized beds and those confined by a solid upper boundary^{13,22}, motions are in the opposite direction to our observations and do not result in axial banding. However, the direction of secondary motion characterized in our fluidized bed is consistent with ‘anomalous’ fluid cell rotations in finite-length cylinders, in which cell-splitting bifurcations have long been known⁴. We find here that the five-band and seven-band states can be induced directly, bypassing the one-band and two-band patterns, if Ω is increased suddenly. In this case, successive bands materialize at both the top and bottom of the bed, at intervals of a few seconds, until the

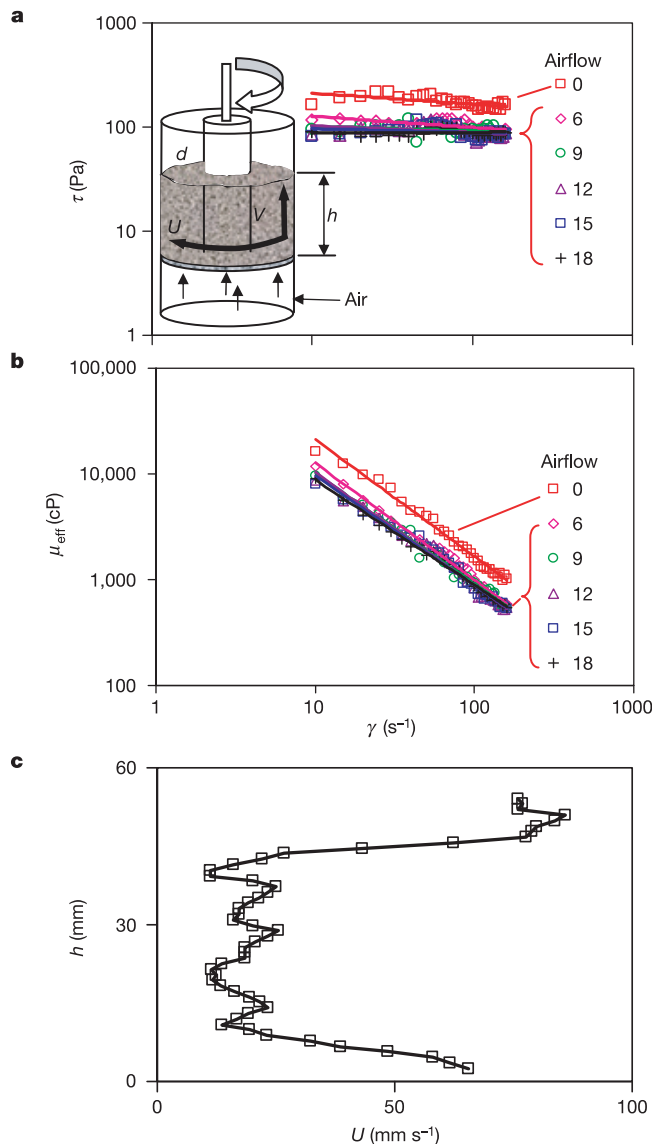


Figure 2 Shear response of fluidized granular bed. **a**, Left: diagram of fluidized Taylor–Couette apparatus. Main plot: effective bed shear stress, τ , determined by torque measurements on the inner cylinder. **b**, Effective bed viscosity, μ_{eff} . Both τ and μ_{eff} are shown as a function of shear rate, γ , for several airflow rates as indicated (measured in litres min⁻¹ at standard temperature and pressure). The solid lines depict strongly shear-thinning behaviour as viscosity decreases in proportion to $\gamma^{-\beta}$, where β ranges from 1 to 1.11. **c**, Azimuthal velocities (U) at outer wall for $Re_{ip} = 50$, obtained by particle-image velocimetry from 1,000 sequential image frames, recorded at 500 frames s⁻¹ with a charge-coupled device camera.

central bed region is filled, matching numerical predictions for fluids by using stress-free endwalls⁵. However, the banding patterns observed do not exhibit any signs of the onset of waviness expected in fluids at higher strains.

To identify relevant physical effects, we conducted experiments with a nearly monodisperse single material (as opposed to our previously bidisperse mixture). We fluidized a bed of the smaller white particles interspersed with coloured, otherwise identical, particles (added for visualization purposes), and flow patterns indicative of circulation cells were observed. Our results with a particle size difference in Fig. 1 indicate that visualization is enhanced by particle size segregation effects, akin to the suspension of polymeric flakes or to pearlescence in fluids^{2,4}, but that the underlying cause of toroidal cell formation remains for both monodisperse and bidisperse beds.

We investigated whether the newly observed instability was centrifugal in nature by modifying our apparatus to accommodate a rotating cylinder with an internal diameter of 64 mm surrounding a 30-mm diameter stationary cylinder. In this geometry, centrifugation has long been known to stabilize, rather than destabilize, vortical flow^{2,4}. Indeed, in our granular case we find that a centrifugation

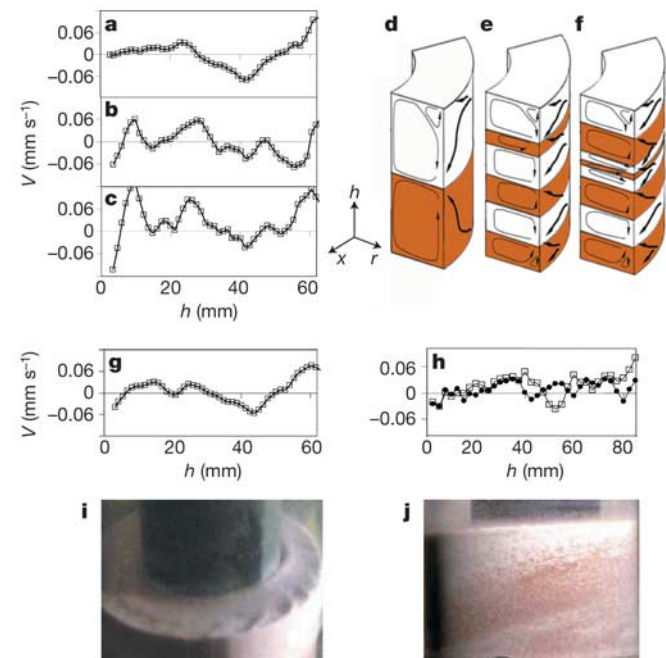


Figure 3 Evidence of Taylor vortices. **a–c**, Axial velocities for segregating system obtained with PIV interrogation of the entire bed height as the transition from one-band to four-band states occurs. Plots correspond to the patterns exemplified in Fig. 1a–c. Sign changes correspond to cell boundaries. **d–f**, Diagrams of both measured two-dimensional wall velocities and inferred three-dimensional circulation patterns within the bulk of the granular bed, corresponding to **a–c**. **g**, Additionally, although not clearly visible to the naked eye, a two-band transient (of which there are signs of development in Fig. 1a) is captured by PIV, representing a state intermediate to those in **a** and **b**. Very brief sign changes such as that developing in **c** at $h \approx 35$ mm correspond to narrow bands (for example Fig. 1c); a fifth band seems about to emerge next at $h \approx 20$ mm. Peak axial magnitudes increase with Ω , indicating that vortices rotate more rapidly. **h**, Quantification of velocity fields for monodisperse material, consisting entirely of glass spheres of diameter $140 \pm 10 \mu\text{m}$. About 10% of the particles are coloured red (but are otherwise mechanically identical), to enable particle tracking by particle image velocimetry. Axial velocities (V) at the outer wall for $\Omega = 450$ r.p.m. (open squares) and 600 r.p.m. (filled circles). **i**, Image of upper, free surface of bed at $Re_{ip} = 61$. Lighting is from the right at a low incidence angle to reveal travelling surface waves near the outer wall. **j**, Side view showing spiral banding pattern at $Re_{ip} = 131$ (550 r.p.m.), induced by airflow maldistribution (see the text).

gal effect is seen in which a large proportion of particles rise up the outer cylinder wall, losing contact with the inner cylinder altogether, but up to rotation speeds of 800 r.p.m. no evidence of toroidal flow in monodisperse blends, or of segregated bands in bidisperse blends, is seen. The impact of small eccentricities or oscillations in the inner or outer cylinder, as well as changes in the aspect ratio of the Couette gap on the effects we report, remain to be fully explored.

To quantify velocities in the states shown in Fig. 1, we use particle image velocimetry (PIV) to measure velocities at exposed surfaces¹⁴. Figure 2c quantifies the azimuthal velocity at the outer wall. A velocity maximum appears at the upper boundary, an observation consistent with a stress-free surface, and a maximum of similar amplitude at the base is evidence that stress due to friction is also relieved near the distributor plate. Axial velocities (V) are also illuminating. These are measured at several points as the flow transitions from one to four bands and are found to fluctuate in sign with h (Fig. 3a–c). Comparison with originating images demonstrates that each sign change marks an interface between downward-flowing small particles and upward-flowing large particles—and by mass conservation we infer opposing motion near the inner cylinder (Fig. 3d–f)⁷. A band is therefore interpreted as a pair of counter-rotating Taylor-like vortices. At the bed base and surface, smaller vortices confined to outer corners seem to be present, perhaps indicative of small, secondary vortices induced in fluids in finite-length cylinders⁴. The PIV technique also reveals a transitional 2-band state at intermediate rotation rates (Fig. 3g).

To determine whether these circulation patterns are inherent in the nature of granular materials in the Couette apparatus or are an artefact of mixture segregation, we replace the granular mixture with a near-monodisperse material possessing a particle size equivalent to the white particles in Fig. 1. Axial velocities (V) are shown for two typical cases in Fig. 3h. For relatively low shear rates, particle velocities make a transition from upward to downward motions near the axial midpoint. At higher shear rates, axial velocities experience several changes in sign with increasing h , in a similar way to the granular mixture. Azimuthal velocities at the outer wall (not shown) are also similar to those observed for the granular mixture (see Fig. 2c). In general, the PIV results for the single material are consistent with those observed for the granular mixture.

As for fluid Taylor vortices in finite-length cylinders, Ekman pumping at the endwalls could be significant. Although particles at both the free surface (or lower boundary) and those in the bed interior are subject to centrifugal forces, associated differences in stress cause either the interior or endwall particles to move radially more rapidly, permitting recirculation zones to be initiated at the endwalls of the cylinder. In fluids this leads both to characteristic Ekman spiral patterns at the free surface and to recirculation cells that can initially fill the entire annulus and meet near the cylinder mid-plane^{2,4}. The interaction between Ekman cells and the Taylor instability is, in general, complex. As Re_i increases, radial pressure gradients grow, typically forcing the Ekman cells into the outer corners at the endwalls and inducing a pair of Taylor vortices in the main flow. This fluid process is consistent with current observations. We observe travelling surface waves resembling Ekman spirals as precursors to subsurface banding (Fig. 3i). In contrast to cylinders with no-slip boundaries, we conjecture that, here, higher azimuthal velocities at the stress-free endwalls cause Ekman vortices to rotate outwards at the cylinder boundaries. Final evidence of the vortex nature of the banding is seen when an axial flow perturbation, prompted by differential injection of air at a single point on the distribution plate, causes bands to spiral (Fig. 3j)—an effect first described by Taylor for fluids¹.

Evidently, fluidization reduces the shear stresses in the granular bed and permits the formation of Taylor-like vortices, albeit with two clear differences from those in fluids. We see no sign that increasing strain is alleviated by the usual secondary instabilities,

even at Reynolds numbers far above those yielding wavy vortices in fluids. Instead, the particle vortices (associated with dissipative collisions^{23,24}) undergo unusual and unexpected stepwise reductions in axial wavelength, λ . Apparently, grains deprived of a solidifying mechanism (that is, stress chains) dissipate internal stresses intrinsically differently from fluids, even in flows that look qualitatively similar to those of their fluid cousins. In particular, rather than producing fixed-wavelength vortices that develop a hierarchy of instabilities with multiple length scales as applied strain rate increases, grains seem to bypass this mechanism and self-select a single new wavelength to dissipate stress. The second new feature is the segregation of large grains into alternate co-rotating vortices. Although segregation by itself is not altogether unexpected in this system^{20,25}, the exact cause and mechanism by which particle species associate with particular vortices are not known. Moreover, it is curious that the segregational flux is apparently large enough to overcome gravitational and fluid forces²¹ tending to produce the small-above-large state of Fig. 1a.

Spontaneous development of a stepwise reduction in length scales and the tenacity of consequent segregation in this prototypical shearing apparatus has so far gone unreported^{26–28}. Yet given the ubiquity of sheared grains, it seems plausible that this analogy to a most instructive hydrodynamic instability is pervasive. The precedent for exploiting the Taylor instability in reacting fluids has been set, and applications such as the blending of pharmaceutical ingredients are ripe²⁹. In particular, our results indicate that shear flows in many granular systems, for example rotating crustal blocks within seismic faults¹⁰, might also evolve discretely, potentially with associated mixing–segregation transitions. It remains to be seen whether the final, mixed state we observe here indicates the onset of other, turbulent-like granular regimes. □

Received 10 March; accepted 30 July 2004; doi:10.1038/nature02901.

1. Taylor, G. I. Stability of a viscous liquid contained between two rotating cylinders. *Phil. Trans. R. Soc. Lond. A* **223**, 289–343 (1923).
2. Andereck, D., Liu, S. S. & Swinney, H. L. Flow regimes in a circular Couette system with independently rotating cylinders. *J. Fluid Mech.* **164**, 155–183 (1986).
3. Mullin, T. & Benjamin, T. B. Transition to oscillatory motion in the Taylor experiment. *Nature* **288**, 567–569 (1980).
4. Mullin, T. Mutations of steady cellular flows in the Taylor experiment. *J. Fluid Mech.* **121**, 207–218 (1982).
5. Abshagen, J., Meincke, O., Pfister, G., Cliffe, K. A. & Mullin, T. Transient dynamics at the onset of Taylor vortices. *J. Fluid Mech.* **476**, 335–343 (2003).
6. Mueth, D. *et al.* Signatures of granular microstructure in dense shear flows. *Nature* **406**, 385–389 (2000).
7. Forterre, Y. & Pouliquen, O. Longitudinal vortices in granular flows. *Phys. Rev. Lett.* **86**, 5886–5889 (2001).
8. Mobius, M. E., Lauderdale, B. E., Nagel, S. R. & Jaeger, H. M. Size separation of granular particles. *Nature* **414**, 270 (2001).
9. Burtally, N., King, P. J. & Swift, N. R. Spontaneous air-driven separation in vertically vibrated fine granular mixtures. *Science* **295**, 1877–1879 (2002).
10. Scott, D. R. Seismicity and stress rotation in a granular model of the brittle crust. *Nature* **381**, 592–595 (1996).
11. Bridgwater, J. The dynamics of granular materials - towards grasping the fundamentals. *Granular Matter* **4**, 175–181 (2003).
12. Michaels, J. N. Toward rational design of powder processes. *Powder Technol.* **138**, 1–6 (2003).
13. Savage, S. B. & Sayed, M. Stresses developed by dry cohesionless granular materials sheared in an annular shear cell. *J. Fluid Mech.* **142**, 391–430 (1984).
14. Wereley, S. T. & Lueptow, R. M. Spatio-temporal character of non-wavy and wavy Taylor–Couette flow. *J. Fluid Mech.* **364**, 59–80 (1998).
15. Losert, W., Bocquet, L., Lubensky, T. & Gollub, J. Particle dynamics in sheared granular matter. *Phys. Rev. Lett.* **85**, 1428–1431 (2000).
16. Medlin, J., Wong, H.-W. & Jackson, R. Fluid mechanical description of fluidized beds: Convective instabilities in bounded beds. *Indust. Eng. Chem. Fundam.* **13**, 247–259 (1974).
17. Apicella, E., D'Amore, M., Tardos, G. & Mauri, R. Onset of instability in sheared gas fluidized beds. *AIChE J.* **43**, 1362–1373 (1997).
18. Wightman, C., Muzzio, F. J. & Wilder, J. A quantitative image analysis method for characterizing mixtures of granular materials. *Powder Technol.* **89**, 165–176 (1996).
19. Wilkes, J. O. *Fluid Mechanics for Chemical Engineers* (Prentice-Hall, Upper Saddle River, 1999).
20. Silva, S. R. D., Dyrøy, A. & Enstad, G. G. in *IUTAM Symposium on Segregation in Granular Flows* (eds Rosato, A. D. & Blackmore, D. L.) 11–30 (Kluwer Academic, Dordrecht, 1999).
21. Djeridi, H., Fave, J.-F., Billard, J.-Y. & Fruman, D. H. Bubble capture and migration in Couette–Taylor flow. *Expts Fluids* **26**, 233–239 (1999).
22. Khosropour, R., Zirinsky, J., Pak, H. K. & Behringer, R. P. Convection and size segregation in a Couette flow of granular material. *Phys. Rev. E* **56**, 4467–4473 (1997).
23. Hirschfeld, D. & Rapaport, D. C. Molecular dynamics simulation of Taylor–Couette vortex formation. *Phys. Rev. Lett.* **80**, 5337–5341 (1998).

24. Conway, S. & Glasser, B. Density waves and coherent structures in granular Couette flows. *Phys. Fluids* **16**, 509–529 (2004).
25. Ottino, J. M. & Khakhar, D. V. Mixing and segregation of granular materials. *Annu. Rev. Fluid Mech.* **32**, 55–91 (2000).
26. Bridgewater, J. Fundamental powder mixing mechanisms. *Powder Technol.* **15**, 215–236 (1976).
27. Joseph, D. D., Wang, J., Bai, R., Yang, B. H. & Hu, H. H. Particle motion in a liquid film rimming the inside of a partially filled rotating cylinder. *J. Fluid Mech.* **496**, 139–163 (2003).
28. Matson, W. R., Ackerson, B. J. & Tong, P. Pattern formation in a rotating suspension of non-Brownian settling particles. *Phys. Rev. E* **67**, 050301 (2003).
29. Shinbrot, T. & Muzzio, F. Nonequilibrium patterns in granular mixing and segregation. *Physics Today* **53**, 25–30 (2000).

Acknowledgements We thank K. LaMarche, D. Brain, S. Shah, M. Ozbas, M. Clark, A. Alexander and D. Trinkle for assistance. This work was partly supported by NSF, ACS-PRF and NASA. S.L.C. thanks Merck & Co. for financial support during an educational leave of absence.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.J.G. (bglasser@rutgers.edu).

Complete photo-fragmentation of the deuterium molecule

T. Weber^{1,2,3}, A. O. Czasch¹, O. Jagutzki¹, A. K. Müller¹, V. Mergel¹, A. Kheifets⁴, E. Rothenberg², G. Meigs², M. H. Prior², S. Daveau², A. Landers⁵, C. L. Cocke³, T. Osipov³, R. Diez Muiño⁶, H. Schmidt-Böcking¹ & R. Dörner¹

¹Institut für Kernphysik, Universität Frankfurt, D 60486 Frankfurt, Germany

²Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

³Department of Physics, Kansas State University, Manhattan, Kansas 66506, USA

⁴Research School of Physical Sciences and Engineering, Australian National University, Canberra ACT 0200, Australia

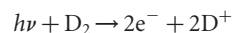
⁵206 Allison Laboratory, Auburn University, Alabama 36849 5311, USA

⁶Donostia International Physics Center and Unidad de Física de Materiales, 20 018 Donostia-San Sebastian, Spain

All properties of molecules—from binding and excitation energies to their geometry—are determined by the highly correlated initial-state wavefunction of the electrons and nuclei. Details of these correlations can be revealed by studying the break-up of these systems into their constituents. The fragmentation might be initiated by the absorption of a single photon^{1–6}, by collision with a charged particle^{7,8} or by exposure to a strong laser pulse^{9,10}:

if the interaction causing the excitation is sufficiently understood, the fragmentation process can then be used as a tool to investigate the bound initial state^{11,12}. The interaction and resulting fragment motions therefore pose formidable challenges to quantum theory^{13–15}. Here we report the coincident measurement of the momenta of both nuclei and both electrons from the single-photon-induced fragmentation of the deuterium molecule. The results reveal that the correlated motion of the electrons is strongly dependent on the inter-nuclear separation in the molecular ground state at the instant of photon absorption.

Small systems of Coulomb interacting particles such as the helium atom or the hydrogen molecule have been models for quantum theory since its early days. However, despite 80 years of theoretical attention, nearly exact calculations for such systems are available only for bound states. On the experimental side, tests of these calculations are based largely on level energies or single-particle momentum distributions. Very promising and challenging new classes of experiments are those that achieve a complete description of the outcome after excitation of the ground state to an unbound continuum. The momenta—that is, the set of vectors—of all fragments of the break-up of an atom or molecule can be measured in coincidence with high precision by using state-of-the-art imaging and timing techniques¹⁶. These asymptotic many-particle momentum distributions are determined by the interaction inducing the fragmentation, the bound initial state from which it emerged, and interactions between the outgoing particles. Thus, experimentalists find it useful to keep the interaction process as simple as possible and to choose a geometry in which final state interactions are negligible. In the present study we used the absorption of a single photon to fragment the deuterium molecule:



Because of the heavy masses of the nuclei, their initial motion in the continuum can be assumed to be the same as in the ground state at the instant of the electronic transition (Born–Oppenheimer approximation). Once the electrons have left the system, the motion of the nuclei is determined solely by their Coulomb repulsion; they accelerate to a kinetic energy release (KER) that corresponds to the Coulomb potential associated with their initial separation. Quantum mechanically, one maps the nuclear vibrational wavefunction onto the Coulomb potential to yield a KER spectrum. Inverting this process determines the squared nuclear vibrational wavefunction from the measured KER spectrum¹⁷. Furthermore, by selecting events that occur within a fixed subregion in the KER spectrum,

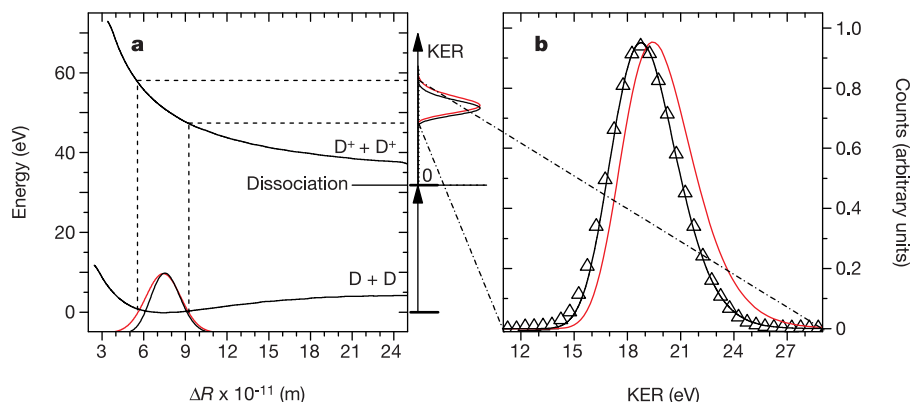


Figure 1 Mapping of the ground-state distribution of inter-nuclear distances to the kinetic energy in the final state. Determination of the inter-nuclear distance ΔR in the ground state of D_2 by measuring the KER after double ionization with linear polarized light of 75.5 eV. In panel **a** the squared ground-state wavefunction is mirrored at the repulsive $\text{D}^+ + \text{D}^+$

potential curve onto the KER (shown as an enlargement in panel **b**): the red curves show the prediction starting from a harmonic oscillator potential; the black curves show the same from a Morse potential. The open triangles are the experimental measurements (the standard deviation of the mean value is smaller than the size of the triangles).

molecules are sampled whose corresponding inter-nuclear distance is defined much more precisely than the full extent of the initial nuclear wavefunction. This allows us to show how the electronic continuum momentum distribution depends on the inter-nuclear separation in the molecule and its orientation with respect to the photon polarization.

Multi-particle coincidence experiments have become possible with the advent of modern imaging techniques based on micro-channel plate multipliers combined with high-resolution time-of-flight measurements^{18–21}. In brief, inside our momentum spectrometer, a supersonic jet of D₂ gas was crossed with the linear polarized photon beam from the Lawrence Berkeley National Laboratory Advanced Light Source (D₂ provides higher target density than a comparable H₂ gas jet, and data are less contaminated by random coincidences from background water). The electrons and ions created in the intersection of the photons with the jet are guided by a combination of homogeneous electric and parallel magnetic fields onto channel plate detectors. For each particle, the

position of impact and the time of flight to the detector are registered. From these values and the geometry and fields, the momentum vectors of all particles from each fragmentation event can be calculated. The validity and sensitivity of the mapping of the ground-state distribution of inter-nuclear distances to the KER is shown in Fig. 1. The red line shows the squared nuclear wavefunction obtained when approximating a D₂ potential curve by a harmonic oscillator with the correct vibration frequency and equilibrium distance; the black curve shows the calculation with

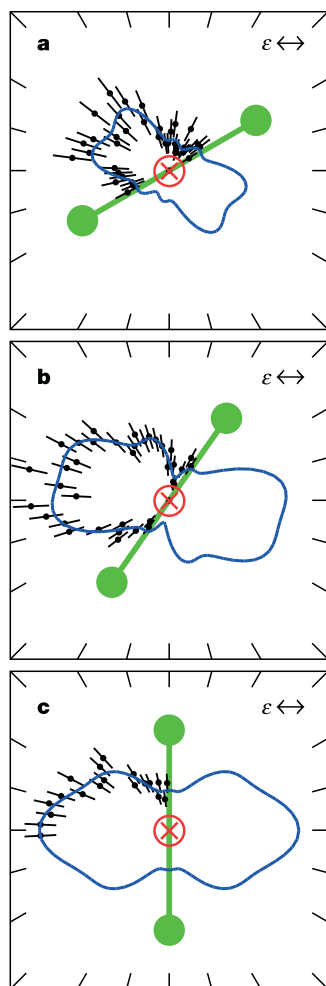


Figure 2 Angular distribution in the non-coplanar frame as a function of molecular orientation. Shown is the angular distribution of one electron (black dots with error bars indicating the standard deviation of the mean value) in the plane of the molecular axis (green barbell) and the electric field vector of the linear polarized light ε (horizontal double arrow) for a photon energy of 75.5 eV. In panel **a** the angle between the molecule and the polarization axis is 30°, whereas it is 55° in **b** and 90° in **c**. The other electron moves orthogonally out of the plane towards the observer (the red circled cross). Each electron has 12.25 eV energy. Because any half-plane is indistinguishable from its inverted complement, we place all data into one half-plane even though we collect ions and electrons in all 4π of laboratory space. The blue solid lines show a fit with spherical harmonics ($l \in [1,4]$, $m \in [0,1]$).

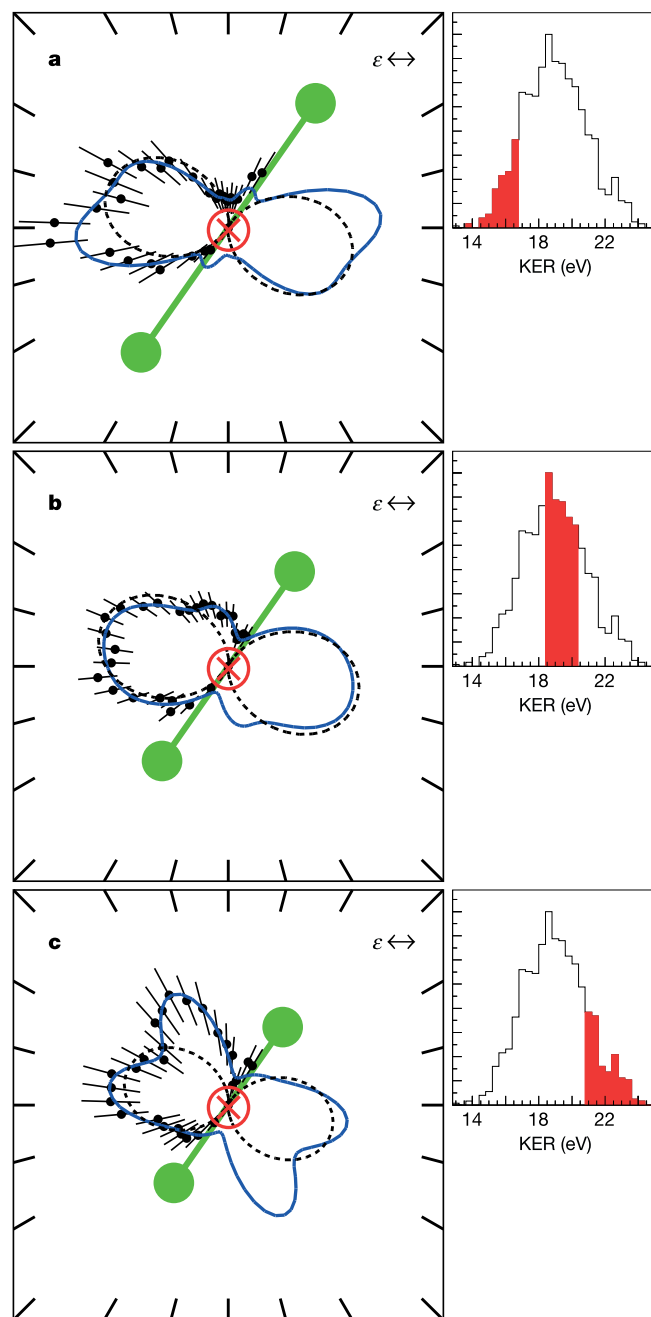


Figure 3 Angular distribution in the non-coplanar frame as a function of KER. Details are as in Fig. 2, but here the molecular axis is fixed at 55° as indicated while the KER varies between panels **a**, **b** and **c** as shown in the insets to the right. High KER corresponds to small inter-nuclear distances ΔR at the instant of photon absorption as described in Fig. 1 (indicated by the varying length of the green 'barbell' in each panel). The blue solid lines show a fit with spherical harmonics ($l \in [1,4]$, $m \in [0,1]$). The dashed black lines represent the result from a single centre expansion of the molecular ground state and a CCC expansion of the final two-electron continuum.

an improved wavefunction obtained from a much more accurate Morse potential.

In Fig. 2 we illustrate the dependence of angular distributions of coincident electrons on the molecular orientation. In this figure the molecular axis, the light polarization axis ε and one of the electrons are restricted to one plane. The momentum vector of the second electron is fixed perpendicular to that plane, pointing towards the observer. By keeping the angle between the two electrons constant at 90° , we expect minimal influence and variation of the electron–electron interaction in the final state. For other geometries we find that the angular distributions are almost completely dominated by the interplay between electron–electron repulsion, angular momentum and parity selection rules^{14,22}.

An atomic photo-ionization cross-section can be described by a dipole distribution with its symmetry axis in the direction of the polarization of the incoming photon. In contrast, Fig. 2, for the molecular case, shows a strong dependence of the angular distribution of the electrons on the molecular orientation, showing the importance of this new internal reference axis. Whereas the light field attempts to drive electrons towards a dipole pattern keyed to its polarization axis, the two-body Coulomb potential of the molecule tends to favour electrons escaping perpendicular to its axis. Calculations with a correlated final state that includes all two-body interactions (except the inter-nuclear repulsion) contradict this behaviour (see ref. 23). Furthermore, the distribution is not describable by a pure dipole shape; it shows additional small lobes (see Fig. 2b, c) indicating that higher-angular-momentum components are present. A possible reason for the preferred emission perpendicular to the molecular axis could be the ground-state electron momentum distribution in the molecule. Because the electronic wavefunction is elongated along the inter-nuclear axis in configuration space, the momentum space wavefunction—that is, the Fourier transform—forms a peak perpendicular to the molecular axis. The photo-ionization probes the momentum-space wavefunction; indeed, the photo-ionization matrix element in the high-energy limits corresponds to the Fourier transform of the initial state²³. In contrast, a description of the ionization process in terms of diffraction of an outgoing electron by the two centres of the potential can yield similar complex angular distribution patterns. From this one might expect that these emission patterns of the photo double ionization shown in Fig. 2 can be described by the single ionization of a D_2^+ ion. Despite these seemingly reasonable qualitative ideas, we show below that this approach is not appropriate.

A key result of our experiment is shown in Fig. 3. The geometry is the same as in Fig. 2, but here the plots are made for selected regions in the KER spectrum. For the smallest KER (that is, the largest separation), the angular distribution resembles a helium-like dipole pattern, in that it is aligned mainly along the polarization axis (see Fig. 3a). For the smallest inter-nuclear separation (Fig. 3c) the emission pattern is essentially orthogonal to the molecular axis and the distribution changes from a dipole to a four-lobed pattern, indicating that higher-angular-momentum components are involved. What is the physical origin of these observations? First, we exclude interference modulations from simple two-centre diffraction, because here the wavelengths of the electrons are four to six times the inter-nuclear separation. Second, whereas a multiple scattering of the photoelectron wave could lead to a variation of the angular distributions, as is predicted and observed for K-shell ionization of the CO molecule for instance²⁴, investigations on H_2 , applying the method described in ref. 25, show that this effect, as a function of the inter-nuclear separation ΔR , is rather small. This is because the protons are relatively weak scattering centres and the long wavelength of the photoelectrons would require long paths within the molecular potential. A remaining possibility might be found in the initial-state electronic wavefunction.

A calculation based on an appropriate two-electron equilibrium

initial state is included in Fig. 3b. The corresponding experimental data show a mixture of patterns similar to the dipole distribution of Fig. 3a and the four-lobed structure of Fig. 3c. A simple model^{26,27}, in which a pair of photo-ionization amplitudes f_Σ and f_Π are introduced for light polarization parallel (Σ transition) and perpendicular (Π transition) to the molecular axis ionizing the molecule, is shown as the dashed black line. To evaluate the amplitudes f_Σ and f_Π we used a single-centre expansion of the H_2 ground state²⁸ and a convergent close-coupling (CCC) expansion of the final two-electron state in the field of a point-like charge $Z = 2$ (ref. 29). This calculation, neglecting two-centre electron–nuclei interaction in the final state, yields only the dipole pattern. The difference between this result and the observations reveals the complex diffraction of the outgoing electron wave from a highly correlated two-electron initial state. Although they reproduce a decrease in magnitude, the CCC calculations corresponding to the kinematics of Fig. 3a, c based on single-centre Slater-type orbitals for different inter-nuclear distances³⁰ (shown also as dashed black lines) do not differ greatly from the result shown for the equilibrium state in Fig. 3b. Apparently this kind of initial-state function is not as sensitive to small changes in ΔR as the experimental data seem to require.

Thus we see complex structures in the electronic angular distribution that depend strongly on the molecular orientation and the inter-nuclear separation. The angular distributions, apparently highly influenced by an appropriate initial-state wavefunction, diffraction and electron–electron correlation, show behaviour that is both unexpected and not yet understood. Our results are highly sensitive and direct tests of the initial-state wavefunction and its correlation effects. An intricate calculation to address our observations is highly desirable. A complete treatment of the break-up of this fundamental molecule would mark a significant step towards understanding the quantum dynamics of many-particle systems, a subject central to most physical and chemical processes. □

Received 18 May; accepted 13 July 2004; doi:10.1038/nature02839.

- Kossmann, H., Schwarzkopf, O., Kämmerling, B. & Schmidt, V. Unexpected behavior of double photoionization in H_2 . *Phys. Rev. Lett.* **63**, 2040–2043 (1989).
- Dujardin, G., Besnard, M. J., Hellner, L. & Malinovich, Y. Double photoionization of H_2 : An experimental test of electronic-correlation models in molecules. *Phys. Rev. Lett.* **35**, 5012–5019 (1987).
- Reddish, T. J., Wightman, J. P., MacDonald, M. A. & Cvejanovic, S. Triple differential cross section measurements for double photoionization of D_2 . *Phys. Rev. Lett.* **79**, 2438–2441 (1997).
- Wightman, J., Cvejanovic, S. & Reddish, T. J. ($\gamma, 2e$) cross section measurements of D_2 and He. *J. Phys. B* **31**, 1753–1764 (1998).
- Secombe, D. P. et al. Photodouble ionization differential cross sections for D_2 with various electron energy sharing conditions. *J. Phys. B* **35**, 3767–3780 (2002).
- Dörner, R. et al. Double photoionization of spatially aligned D_2 . *Phys. Rev. Lett.* **81**, 5776–5779 (1998).
- Afaneh, F. et al. Must saddle point electrons always ride on the saddle? *J. Phys. B* **35**, L229–L235 (2002).
- Wood, R. M., Edwards, A. K. & Steuer, M. F. Dissociative ionization of H_2 and D_2 produced by bombardment with fast He^+ ions. *Phys. Rev. A* **4**, 1433–1437 (1977).
- Staudte, A. et al. Observation of a nearly isotropic, high-energy Coulomb explosion group in the fragmentation of D_2 by short laser pulses. *Phys. Rev. A* **65**, 020703–1–020703–4 (2002).
- Rotke, H. et al. Coincident fragment detection in strong field photoionization and dissociation of H_2 . *Phys. Rev. Lett.* **89**, 013001–1–013001–4 (2002).
- Moshhammer, R. et al. The dynamics of target ionization by fast highly charged projectiles. *Nucl. Instr. Meth. Phys. Res. B* **107**, 62–66 (1996).
- Levin, V. G., Neudatchin, V. G., Pavlitchankov, A. V. & Smirnov, Yu. F. A study of the electron correlations in the H_2 molecule using the double photoionisation process ($\gamma, 2e$). *J. Phys. B* **17**, 1525–1536 (1984).
- Rescigno, T. N., Baertschy, M., Isaacs, W. A. & McCurdy, C. W. Collisional breakup in a quantum system of three charged particles. *Science* **286**, 2474–2479 (1999).
- Weber, T. et al. Fully differential cross sections for photo-double-ionization of D_2 . *Phys. Rev. Lett.* **92**, 163001–1–163001–4 (2004).
- Briggs, J. & Schmidt, V. Differential cross section for photo-double-ionization of the helium atom. *J. Phys. B* **33**, R1–R48 (2000).
- Dörner, R. et al. Photo-double-ionization of He: Fully differential and absolute electronic and ionic momentum distributions. *Phys. Rev. A* **57**, 1074–1090 (1998).
- Le Rouzo, H. Double photoionization of molecular hydrogen: A theoretical study including the nuclear dissociation. *Phys. Rev. A* **37**, 1512–1523 (1988).
- Dörner, R. et al. Cold target recoil ion momentum spectroscopy: a ‘momentum microscope’ to view atomic collision dynamics. *Phys. Rep.* **330**, 95–192 (2000).
- Ullrich, J. et al. Recoil-ion and electron momentum spectroscopy: reaction-microscopes. *Rep. Prog. Phys.* **66**, 1463–1545 (2003).

20. Weber, T. *et al.* Correlated electron emission in multiphoton double ionization. *Nature* **405**, 658–661 (2000).
21. Schulz, M. *et al.* Three-dimensional imaging of atomic four-body processes. *Nature* **422**, 48–50 (2003).
22. Walter, M. & Briggs, J. S. Selection rules and isotope effects in the full fragmentation of the hydrogen molecule. *Phys. Rev. Lett.* **85**, 1630–1633 (2000).
23. Walter, M. & Briggs, J. S. Photo-double ionization of molecular hydrogen. *J. Phys. B* **32**, 2487–2501 (1999).
24. Weber, T. *et al.* Auger electron emission from fixed-in-space CO. *Phys. Rev. Lett.* **90**, 153003–1–153003–4 (2003).
25. Diez Muino, R., Rolles, D., de Abajo, F. J. G., Fadley, C. S. & Hove, M. A. V. Angular distribution of the electrons photoemitted from core levels of oriented diatomic molecules: multiple scattering theory in non-spherical potentials. *J. Phys. B* **35**, L359–L365 (2002).
26. Feagin, J. M. A helium-like description of molecular hydrogen photo-double ionization. *J. Phys. B* **31**, L729–L736 (1998).
27. Reddish, T. J. & Feagin, J. M. Photo double ionization of molecular deuterium. *J. Phys. B* **32**, 2473–2486 (1999).
28. Joy, H. W. & Parr, R. G. A one-center wave function for the hydrogen molecule. *J. Chem. Phys.* **28**, 448–453 (1958).
29. Kheifets, A. S. & Bray, I. Application of the CCC method to the calculation of helium-photoionization triply differential cross sections. *J. Phys. B* **31**, L447–L453 (1998).
30. Hayes, E. F. Accurate single-center expansions using Slater type orbitals: hydrogen molecule. *J. Chem. Phys.* **46**, 4004–4008 (1967).

Acknowledgements We thank Roentdek GmbH (www.Roentdek.com) for support with detectors, and acknowledge helpful discussion with colleagues M. Walter, J. Briggs, J. Feagin, T. Reddish and V. Schmidt. This work was supported by the Deutsche Forschungsgemeinschaft, the Bundesministerium für Bildung und Forschung, and the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, Office of Science, US Department of Energy (DOE). T.W. thanks Graduiertenförderung des Landes Hessen, the Alexander von Humboldt Stiftung and the Herrmann Willkomm Stiftung for financial support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.D. (doerner@hsb.uni-frankfurt.de).

Ecosystem carbon storage in arctic tundra reduced by long-term nutrient fertilization

Michelle C. Mack^{1*}, Edward A. G. Schuur^{1*}, M. Sydonia Bret-Harte², Gaius R. Shaver³ & F. Stuart Chapin III²

¹Department of Botany, University of Florida, Gainesville, Florida 32611, USA

²Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, Alaska 99775, USA

³The Ecosystems Center, Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA

* These authors contributed equally to this work

Global warming is predicted to be most pronounced at high latitudes, and observational evidence over the past 25 years suggests that this warming is already under way¹. One-third of the global soil carbon pool is stored in northern latitudes², so there is considerable interest in understanding how the carbon balance of northern ecosystems will respond to climate warming^{3,4}. Observations of controls over plant productivity in tundra and boreal ecosystems^{5,6} have been used to build a conceptual model of response to warming, where warmer soils and increased decomposition of plant litter increase nutrient availability, which, in turn, stimulates plant production and increases ecosystem carbon storage^{6,7}. Here we present the results of a long-term fertilization experiment in Alaskan tundra, in which increased nutrient availability caused a net ecosystem loss of almost 2,000 grams of carbon per square meter over 20 years. We found that annual aboveground plant production doubled during the experiment. Losses of carbon and nitrogen from

deep soil layers, however, were substantial and more than offset the increased carbon and nitrogen storage in plant biomass and litter. Our study suggests that projected release of soil nutrients associated with high-latitude warming may further amplify carbon release from soils, causing a net loss of ecosystem carbon and a positive feedback to climate warming.

The effects of climate warming on ecosystem carbon (C) storage remain uncertain. Despite the low temperatures at high latitudes, C storage in tundra and boreal ecosystems is thought to be constrained ultimately by carbon–nutrient interactions because plant production is usually nitrogen (N)-limited^{6,7}. As soils warm in response to climate change, nutrient mineralization from soil organic matter is expected to increase^{8,9}, which should, in turn, increase plant production. Total ecosystem C storage, however, depends on the balance between production and decomposition, and the relationship between nutrient availability and decomposition remains uncertain.

In ecosystems at lower latitudes, natural and manipulated nutrient concentrations have had a positive, a negative, or no effect on the decomposition of litter and soil organic C (SOC)^{10–13}. This variable response probably reflects ecosystem differences in form and quality of litter and SOC, but the regulatory mechanism for this is poorly understood¹³. High-latitude ecosystems are unusual because they store a larger proportion of total ecosystem C in soil compared with temperate and tropical ecosystems¹⁴. In arctic tundra, as much as 90% of the total ecosystem C resides in organic horizons and frozen mineral soils¹⁵. Thus, the response of SOC to changes in nutrient availability will play a critical role in determining net ecosystem C balance in a changing climate.

Previous results from nutrient manipulations suggested that increased nutrient availability should increase the total C storage in tundra ecosystems^{2,9,15,16}. Nutrient addition greatly increases C stored aboveground by stimulating plant productivity and by shifting species composition from slow-growing species to more productive shrubs that accumulate C in long-lived woody biomass^{4,17–19}. In addition, leaf, root and stem litter from shrubs decomposes more slowly than the graminoid-dominated litter they replace⁹, so conversion to shrub tundra was thought to slow decomposition and increase ecosystem C accumulation¹⁹. However, these inferences were based on aboveground and surface soil measurements only. The lack of soil-profile measurements reflects the expectation that the large heterogeneous belowground C pool

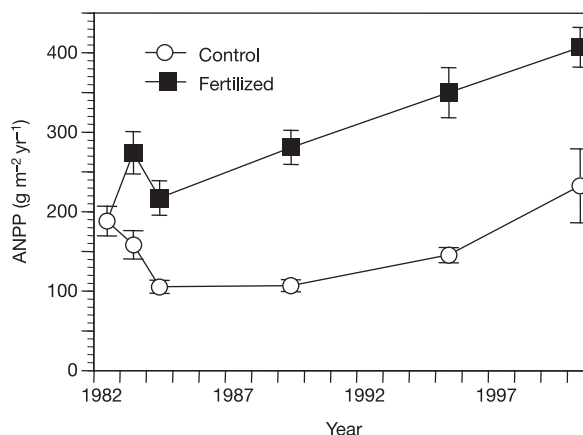


Figure 1 Effect of fertilization on vascular plant aboveground net primary production (ANPP) in tundra. Fertilized plots in moist acidic tundra near Toolik Lake, Alaska, have received 10 g N m⁻² yr⁻¹ and 5 g P m⁻² yr⁻¹ since 1981. Values are means (±1 standard error, s.e.); means from 1982–95 are reported in ref. 19; the year-2000 data are from this study (*n* = 4). Components of ANPP (new leaves and reproductive parts, new stems and secondary growth) are shown in Supplementary Fig. 1.

would respond to changes in plant inputs too slowly to be detected in short-term experimental manipulations.

To investigate the effects of nutrient availability on whole-ecosystem C balance, we examined C and N pools in a long-term fertilization experiment at the arctic Long-Term Ecological Research site near Toolik Lake, Alaska. To our knowledge, this is the longest-running nutrient-addition experiment in arctic tundra. Fertilized plots in moist acidic tundra (MAT) have received 10 g N and 5 g P m⁻² year⁻¹ since 1981 (ref. 19). This is approximately 5 to 8 times the annual soil N uptake requirement for aboveground production in MAT, and similar to the N uptake requirement in nearby shrub tundra characteristic of warmer sites^{5,20}. Two decades of fertilization greatly increased aboveground net primary productivity (ANPP; Fig. 1); over this time about 1,500 g m⁻² of additional C entered fertilized plots as ANPP, which had shifted from graminoid tundra dominated by the tussock-forming sedge, *Eriophorum vaginatum*, to shrub tundra dominated by *Betula nana*¹⁹. Because of the unique long-term nature of this experiment, changes in belowground C pools and total ecosystem C balance are now detectable.

In our experiment, increased nutrient availability had a larger effect on decomposition than on plant production, resulting in a net loss of almost 2,000 g C m⁻² from this ecosystem over 20 yr (Fig. 2a; $P = 0.04$). Carbon storage increased aboveground ($P < 0.001$) because of the accumulation of woody shrub biomass and litter, but this was offset by a larger decrease of C in belowground pools ($P = 0.02$) due to a pronounced decrease in the C contained in deep organic (>5 cm depth) and upper mineral soil layers (Fig. 2b). The decrease in the deep organic layer C pool was the result of a reduction in the thickness of the layer, because neither %C nor bulk density was affected by fertilization (Supplementary Information). In the upper mineral soil, fertilization reduced %C by 50% ($P = 0.04$), whereas the depth to the frozen soil surface and mineral

soil bulk density did not change (Supplementary Information).

Decreased C storage in the fertilized treatment does not appear to be caused by decreased plant production. As expected, ANPP of vascular plants was higher in fertilized plots (Fig. 1). Root C pools were not different between treatments (Fig. 2b), and in a related study, root productivity tended to be higher in fertilized plots²¹. The productivity of mosses and lichens was not measured in this harvest, but their production has been estimated as 25–60 g C m⁻² in unmanipulated MAT^{5,20,22}. Although they were mostly absent from fertilized plots in both 1995 and 2000, the loss of their productivity would be insufficient to offset the large increases in vascular productivity. Because plant production increased total C inputs, the net loss of C from the fertilized plots could only have been caused by accelerated decomposition of C.

Several mechanisms could have contributed to the nutrient-induced acceleration of decomposition. First, nutrient additions could have altered the decomposability of fresh plant litter through changes in the species composition or tissue quality of the plant community. However, the increased C in biomass, litter and surface soils of fertilized plots (1,311 g m⁻²) was similar to increased ANPP inputs over the past 20 yr (~1,500 g m⁻²), indicating relatively little decomposition of the increased litter inputs. Furthermore, *B. nana* leaves, stems and roots are relatively less decomposable than those of the community they replace⁹, making it unlikely that litter decomposed more quickly in the fertilized treatment owing to changes in tissue quality alone. These observations argue against changes in litter decomposability having a major role in increased decomposition in fertilized plots.

Second, the loss of deep-rooted graminoid species from the fertilized plots¹⁹ could have altered environmental controls over root decomposition by changing the depth at which root litter was deposited. Although total root biomass C was not different between treatments, fertilization did shift the distribution of root biomass

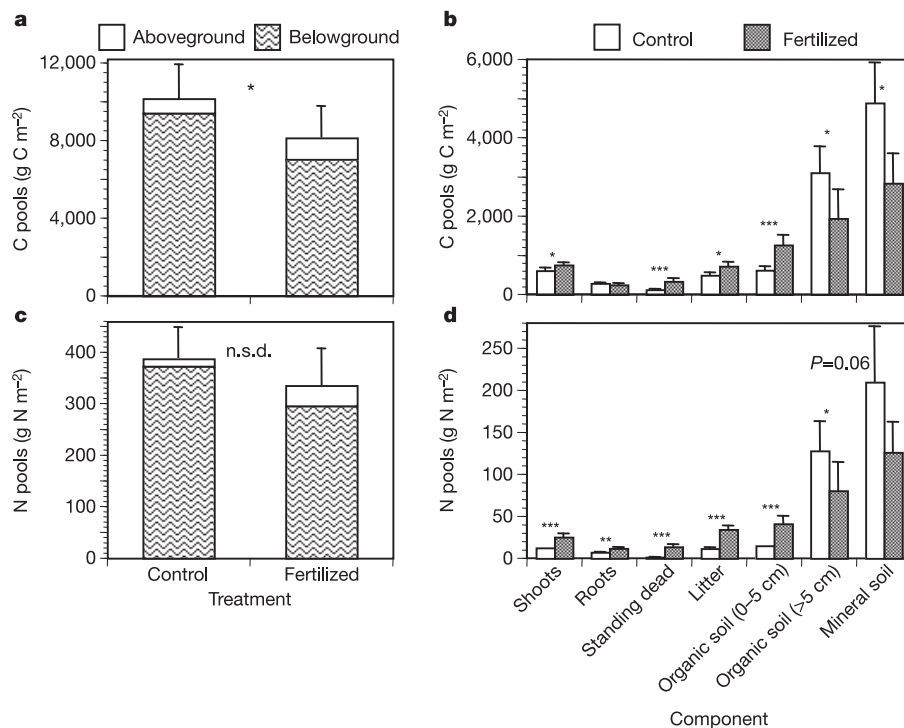


Figure 2 Effects of fertilization on tundra carbon and nitrogen pools after 20 yr of fertilization. **a, c**, Mean (± 1 s.e.) above- and belowground carbon (**a**) and nitrogen (**c**) pools in unmanipulated control and fertilized treatments of moist acidic tundra near Toolik Lake, Alaska. Aboveground pools include shoots, standing dead plant material, and rhizomes. Belowground pools include surface litter, roots, and organic and mineral soil.

b, d, Mean (± 1 s.e.) carbon (**b**) and nitrogen (**d**) pools in plant and soil compartments. Pool treatment means were compared with nested ANOVA ($n = 4$). Means that are significantly different are indicated with asterisks: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. n.s.d., not significantly different.

towards the soil surface. Seventy percent of root biomass was in litter and surface organic soils (0–5 cm) in the fertilized plots, whereas only 30% of root biomass was in these layers in the control plots ($P < 0.05$). Furthermore, there was 50% less root biomass in the mineral soil layer of fertilized plots than in control plots ($P < 0.01$). Because of the large gradients in temperature and moisture in soils overlaying permafrost, this upward shift of root biomass in the fertilized plots may have placed root inputs into a warmer, drier environment where decomposition was more rapid. However, even if root decomposition rates doubled in the fertilized plots, root production ($\sim 34 \text{ g C m}^{-2} \text{ yr}^{-1}$; ref. 21) would constrain the amount of C available for decomposition. At most, increased root decomposition could explain 17% of the C missing from the fertilized plots.

For the remainder of the missing C, we hypothesize that increased nutrient availability stimulated the decomposition of old litter and/or SOC in deep soil layers, leading to loss via mineralization or leaching of dissolved organic C. Other experiments in tundra have also detected increased decomposition rates with increased N availability for fresh leaf litter²³ and for SOC fractions¹³, or have inferred N limitation of decomposer activity²⁴. These results may have been caused directly by nutrient stimulation of decomposer activity, which occurs in systems where litter C quality is low and C:N ratios are high¹². Alternatively, indirect effects of nutrient addition on plant–microbe interactions and soil food-web dynamics may have caused these results. Decreased C:N ratios are expected to shift microbial community composition from fungal to bacterial dominance, resulting in greater energy flow to higher trophic levels and more rapid decomposition²⁵. Regardless of the specific mechanism, the clear result of this study is that increased nutrient availability enhanced decomposition of belowground C pools in deep soil layers more than it increased primary production, leading to a substantial net loss of C from this ecosystem. As an increasing number of ecosystems are subjected to human-driven change in climate and N inputs, the sensitivity of decomposition to nutrient availability may be important for determining ecosystem C balance worldwide.

Even though a total of 200 g N m^{-2} was added to fertilized plots over 20 yr, our results show that fertilization had no net effect on total N pools (Fig. 2c). Increased N storage in aboveground biomass ($P < 0.001$) was balanced by a trend towards decreased N storage belowground ($P = 0.14$). As with C pools, N pools in the fertilized treatment were significantly smaller in deep organic soils and tended to be smaller in mineral soil (Fig. 2d). Concentrations of NH_4^+ and NO_3^- were an order of magnitude higher in the fertilized than in the control plots at all soil depths (Supplementary Information), suggesting that fertilizer N penetrated to deeper soil layers.

Net loss of N from deep soil despite penetration of fertilizer N shows that decomposers in these layers were unable to immobilize N, perhaps owing to C limitation caused by the low soil C:N ratio and the high respiratory demands of cold soils. Furthermore, decreased root biomass at depth probably reduced plant competition for organic and mineral N, leading to increased potential for nitrification and leaching losses of dissolved organic N and nitrate to downslope and aquatic ecosystems²⁶. High concentrations of NO_3^- in the fertilized plots indicate that denitrification may be an important loss pathway as well. These results show that the fertilized ecosystem had almost no capacity for net retention of increased N inputs despite 20 yr of sustained increase in plant production (Fig. 1).

Although it remains to be seen how this response to fertilization will compare with the effects of climate warming on tundra ecosystems, several lines of evidence suggest that a nutrient-mediated positive feedback could occur, amplifying the predicted temperature response of decomposition. First, the amount of N released by warming is likely to be similar in magnitude to our fertilizer addition, ranging from the release of 7 to $9.4 \text{ g N m}^{-2} \text{ yr}^{-1}$

for a 3- to 7-degree increase in mean annual temperature, respectively (see Methods). Second, naturally occurring shrub tundra characteristic of warmer sites has less soil C storage than tussock tundra²⁷ despite its greater nitrogen availability⁸ and ANPP⁷, suggesting that environmental shifts to increased shrub abundance will lead to decreased ecosystem C storage. Finally, we observed the effects of nutrient availability on C storage in the absence of other effects of climate warming, such as increased soil temperature, decreased soil moisture, and increased depth of thaw. Each of these changes is predicted to have a positive effect on decomposition⁷. The observed sensitivity of deep soil C decomposition to increased nutrient availability could amplify these effects and further stimulate C losses from high-latitude systems.

These results from the Toolik Lake fertilization experiment confound predictions of how nutrient-limited northern ecosystems will respond to climate warming. Previous predictions have suggested that warmer soils should stimulate decomposition and increase nutrient availability, which, in turn, will stimulate plant production more than decomposition and increase ecosystem C storage^{6,7}. Our results show the opposite response to increased nutrient availability. Decomposition was stimulated more than plant production, leading to a net loss of C from the ecosystem. Furthermore, the system had no capacity for net retention of increased N inputs. These results suggest that a new conceptual model for the response of tundra to climate warming is required, one in which decomposition is more sensitive than production to changes in nutrient availability and N is susceptible to loss. □

Methods

Site description

The experimental plots were located in moist acidic tundra near Toolik Lake, Alaska, at the Arctic Tundra Long Term Ecological Research site in the northern foothills of the Brooks Range. The soil is a histic pergelic cryaquept, with a ~ 20 cm peat layer over silty mineral soil and permafrost. Mean air temperature during the June–August growing season is $\sim 9.3^\circ\text{C}$ and total precipitation is ~ 180 mm. The vegetation in the experimental plots is similar to moist acidic tundra across the Alaskan North Slope, northern Canada, and eastern Siberia²⁸.

Experimental design

This experiment was established in 1981 as described in detail elsewhere⁴. The design consisted of four replicate blocks of moist acidic tundra; each block contained four $5 \text{ m} \times 20 \text{ m}$ plots. Treatments, including control and fertilization, were randomly assigned. The fertilized treatment has received $10 \text{ g m}^{-2} \text{ N}$ as $\text{NH}_4\text{-NO}_3$ and $5 \text{ g m}^{-2} \text{ P}$ as P_2O_5 each spring immediately following snowmelt since 1981.

Harvest and analyses of ecosystem C and N pools

In late July 2000, we destructively harvested five $20 \text{ cm} \times 20 \text{ cm}$ quadrats from each plot. Quadrats were randomly arrayed along 20-m transects located 1 m from the eastern edge of the plot. Shoots, litter, belowground stems and rhizomes were collected by sawing the organic soil down to the surface of the mineral soil with a knife. Quadrats were returned to the lab and sorted using the methods of previous harvests^{5,19}.

Organic soil was harvested volumetrically from $5 \text{ cm} \times 5 \text{ cm}$ monoliths taken from the side of the hole where the $20 \text{ cm} \times 20 \text{ cm}$ quadrat was removed. Mineral soil was sampled with a 2.5-cm-diameter corer beneath each monolith, from the surface of the mineral soil to the permafrost. Organic soil monoliths were separated into depth increments in the field, including litter (recognizable dead plant material), 0–5 cm organic, 5–15 cm organic, and >15 cm organic. Soil samples were returned to the laboratory, where live roots were removed by hand and sorted into coarse ($>2 \text{ mm}$) and fine ($<2 \text{ mm}$) size fractions. Rock volume and mass were determined and subtracted from soil volume and mass. The remaining soil was subsampled for gravimetric water content and C and N concentration.

All plant and soil samples were dried to a constant mass at 65°C and weighed. As in previous harvests from this experiment^{5,19}, samples from individual quadrats were composited by plot for analyses of C and N content (that is, $n = 4$). Plant shoots, belowground stems, and rhizomes were analysed for C and N concentration on a Perkin-Elmer CHN analyser (Norwalk, Connecticut, USA). Root and soil C and N concentrations were determined on a Carlo Erba CHN analyser (Milan, Italy). Soil %C is reported for bulk soil.

Carbon and N pools were estimated for shoots (aboveground parts, belowground stems and rhizomes), aboveground litter, roots (>5 and $<5 \text{ mm}$ in diameter), and soils for each quadrat by multiplying biomass or soil mass by the plot element concentration for that pool. Total ecosystem C and N pools were estimated by summing all pools. Treatment effects were analysed using nested analyses of variance²⁹, with blocks nested within treatments. Aboveground NPP was estimated as the sum of measured apical and estimated secondary growth¹⁸.

Estimating N release from climate warming

Assuming that (1) annual heterotrophic respiration in unmanipulated tundra is roughly equal to total plant production ($150 \text{ g C m}^{-2} \text{ yr}^{-1}$) as it is in many mature ecosystems, (2) that soil respiration responds to temperature change with a Q_{10} of 2, and (3) the depth-weighted average C:N ratio of SOM is 26 (Supplementary Information), then the projected 3°C temperature increase¹ should result in the mineralization of $7 \text{ g N m}^{-2} \text{ yr}^{-1}$. A 7°C temperature increase should result in the mineralization of $9.4 \text{ g N m}^{-2} \text{ yr}^{-1}$. This is a conservative estimate of temperature stimulation of N release because respiration in soils from cold regions tends to be highly sensitive to temperature, often responding to temperature change with Q_{10} values greater than 2 (ref. 30). Nevertheless, these simple calculations show that our rate of N addition is similar in magnitude to potential N release from climate warming.

Received 4 January; accepted 20 July 2004; doi:10.1038/nature02887.

1. Serreze, M. C. Observational evidence of recent change in the northern high-latitude environment. *Clim. Change* **46**, 159–207 (2000).
2. Gorham, E. Northern peatlands: role in the carbon cycle and probable response to climatic warming. *Ecol. Appl.* **1**, 182–195 (1991).
3. Oechel, W. C. et al. Acclimation of ecosystem CO_2 exchange in the Alaskan Arctic in response to decadal climate warming. *Nature* **406**, 978–981 (2000).
4. Goulden, M. L. et al. Sensitivity of boreal forest carbon balance to soil thaw. *Science* **279**, 214–217 (1997).
5. Chapin, F. S. I., Shaver, G. R., Giblin, A. E., Nadelhoffer, K. J. & Laundre, J. A. Responses of arctic tundra to experimental and observed changes in climate. *Ecology* **76**, 694–711 (1995).
6. Hobbie, S. E., Nadelhoffer, K. J. & Hogberg, P. A synthesis: The role of nutrients as constraints on carbon balances in boreal and arctic regions. *Plant Soil* **242**, 163–170 (2002).
7. Shaver, G. R. et al. Global change and the carbon balance of the ecosystem. *Bioscience* **42**, 433–442 (1992).
8. Nadelhoffer, K. J., Giblin, A. E., Shaver, G. R. & Laundre, J. A. Effects of temperature and substrate quality on element mineralization in six arctic soils. *Ecology* **72**, 242–253 (1990).
9. Hobbie, S. E. Temperature and plant species control over litter decomposition in Alaskan tundra. *Ecol. Monogr.* **66**, 503–522 (1996).
10. Vitousek, P. M. Nutrient cycling and nutrient use efficiency. *Am. Nat.* **119**, 553–572 (1982).
11. Hobbie, S. E., Schimel, J. P., Trumbore, S. E. & Randerson, J. R. A mechanistic understanding of carbon storage and turnover in high-latitude soils. *Glob. Change Biol.* **6**, 196–210 (2000).
12. Berg, N. & Matzner, E. Effect of N deposition on decomposition of plant litter and soil organic matter in forest systems. *Environ. Rev.* **5**, 1–25 (1997).
13. Neff, J. C. et al. Variable effects of nitrogen additions on the stability and turnover of soil carbon. *Nature* **419**, 915–917 (2002).
14. Dixon, R. K. et al. Carbon pools and flux of global forest ecosystems. *Science* **263**, 185–190 (1994).
15. McKane, R. B. et al. Climatic effects on tundra carbon storage inferred from experimental data and a model. *Ecology* **78**, 1170–1187 (1997).
16. Sturm, M., Racine, C. & Tape, K. Climate change: Increasing shrub abundance in the Arctic. *Nature* **411**, 546–547 (2001).
17. Jonasson, S., Michelsen, A., Schmidt, I. K. & Nielsen, E. V. Responses in microbes and plants to changed temperature, nutrient and light regimes in the arctic. *Ecology* **80**, 1828–1843 (1999).
18. Bret-Harte, M. S., Shaver, G. R. & Chapin, F. S. I. Primary and secondary stem growth in arctic shrubs: Implications for community response to environmental change. *J. Ecol.* **90**, 251–267 (2002).
19. Shaver, G. R. et al. Species composition interacts with fertilizer to control long-term change in tundra productivity. *Ecology* **82**, 3163–3181 (2001).
20. Shaver, G. R. & Chapin, F. S. I. Production:biomass relationships and element cycling in constricting arctic vegetation types. *Ecol. Monogr.* **61**, 1–31 (1991).
21. Nadelhoffer, K. J., Johnson, L. C., Laundre, J. A., Giblin, A. E. & Shaver, G. R. Fine root production and nutrient content in wet and moist arctic tundras as influenced by chronic fertilization. *Plant Soil* **242**, 107–113 (2002).
22. Hobbie, S. E. & Chapin, F. S. III The response of tundra plant biomass, aboveground production, nitrogen, and CO_2 flux to experimental warming. *Ecology* **79**, 1526–1544 (1998).
23. Hobbie, S. E. & Gough, L. Litter decomposition in moist acidic and non-acidic tundra with different glacial histories. *Oecologia* **140**, 113–124 (2004).
24. Weintraub, M. N. & Schimel, J. P. Interactions between carbon and nitrogen mineralization and soil organic matter chemistry in arctic tundra soils. *Ecosystems* **6**, 129–143 (2003).
25. Moore, J. C., McCann, K., Setälä, H. & De Ruiter, P. C. Top-down is bottom-up: does predation in the rhizosphere regulate aboveground dynamics? *Ecology* **84**, 846–857 (2003).
26. Giblin, A. E., Nadelhoffer, K. J., Shaver, G. R., Laundre, J. A. & McKerrrow, A. J. Biogeochemical diversity along a riverside toposequence in arctic Alaska. *Ecol. Monogr.* **61**, 415–435 (1991).
27. Michaelson, G. J., Ping, C. L. & Kimble, J. M. Carbon storage and distribution in tundra soils of arctic Alaska. *U.S.A. Arctic Alpine Res.* **28**, 414–424 (1996).
28. Bliss, L. C. & Matveyeva, N. V. in *Arctic Ecosystems in a Changing Climate* (ed. Chapin, F. S. I.) (Academic, San Diego, California, 1992).
29. *Systat for Windows: Statistics Version 7* 7th edn (SYSTAT, Inc., Evanston, Illinois, 1997).
30. Kirschbaum, M. U. F. Will changes in soil organic carbon act as a positive or negative feedback on global warming? *Biogeochemistry* **48**, 21–51 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This research was supported by NSF, NASA, the Arctic LTER programme and an NSF Postdoctoral Fellowship to M.C.M. We thank E. Mitchell, M. Sommerkorn and M. Williams for assistance with the belowground harvest and J. Laundre for laboratory analyses. S. Hobbie and the UF Plant Ecology group provided comments that improved this manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.C.M. (mcmack@ufl.edu).

Environmental predictors of pre-European deforestation on Pacific islands

Barry Rolett¹ & Jared Diamond²

¹Department of Anthropology, University of Hawaii, Honolulu, Hawaii 96822, USA

²Department of Geography, University of California, Los Angeles, California 90095-1524, USA

Some Pacific island societies, such as those of Easter Island and Mangareva, inadvertently contributed to their own collapse by causing massive deforestation^{1–7}. Others retained forest cover and survived^{3,8,9}. How can those fateful differences be explained? Although the answers undoubtedly involve both different cultural responses of peoples and different susceptibilities of environments, how can one determine which environmental factors predispose towards deforestation and which towards replacement of native trees with useful introduced tree species? Here we code European-contact conditions and nine environmental variables for 81 sites on 69 Pacific islands from Yap in the west to Easter in the east, and from Hawaii in the north to New Zealand in the south. We thereby detect statistical decreases in deforestation and/or forest replacement with island rainfall, elevation, area, volcanic ash fallout, Asian dust transport and makatea terrain (uplifted reef), and increases with latitude, age and isolation. Comparative analyses of deforestation therefore lend themselves to much more detailed interpretations than previously possible. These results might be relevant to similar deforestation-associated collapses (for example, Fertile Crescent, Maya and Anasazi) or the lack thereof (Japan and highland New Guinea) elsewhere in the world.

All Pacific islands suitable for agriculture were occupied before European arrival by colonists originating ultimately from Asia, mostly in a wave of Polynesians and their Lapita ancestors from 1200 BC to AD 1200 (refs 2, 10, 11). They cleared land and cut trees, especially for agriculture, timber and fuel. Early European visitors observed that forest cover varied greatly between islands, from totally deforested with almost all original tree species extinct (Easter, Necker and Nihoa) to extensive forests (Samoa, Taveuni and Bismarck Archipelago). Forests seen by early Europeans also varied greatly in composition, from ones still dominated by native species to others whose native species had been largely replaced by introduced species valuable for arboriculture.

From accounts of early European visitors, we coded five-point scales for deforestation (in which 1 represented virtually no deforestation and 5 complete deforestation) and for forest replacement (in which 1 represented virtually no replacement and 5 complete replacement) (see Methods and Supplementary Table S1 for details). Deforestation and replacement proved to be correlated, but they measure different things and the correlation is not tight (Spearman correlation $r = 0.43$, $p < 0.001$, in our data set of 81 entries). For 12 of our 69 islands we coded two different locations on the island (usually windward and leeward coasts) because of very different rainfall values often associated with different degrees of deforestation. Thus, our full data set had $69 + 12 = 81$ entries. We also analysed a reduced data set of 69 entries, which excluded 12 large islands of northern Melanesia and New Zealand, because we wondered whether those 12 islands might be driving some conclusions, but results were similar.

We performed four types of statistical analysis (see Methods) for both data sets, relating our two outcome variables (deforestation and replacement) to nine independent variables discussed below.

The first statistical analysis was the calculation of bivariate Spearman correlations (Supplementary Tables S2–S6), and the corresponding bivariate regression coefficients, among the two outcomes and the nine independent variables. The second analyses were multivariate regression analyses¹², to take account of correlations among the independent variables. The third were multivariate tree models to divide the data, by recursive binary partitioning based on potentially more than one independent variable, into groups each homogenous in outcome value within the group but maximally contrasting in outcome value between groups¹³. Unlike multivariate analysis, this approach seeks sets of explanatory variables that might interact conditionally instead of acting independently. The last analysis was the examination of residuals (actual values minus model-predicted values) from the best-fit multiple regression and tree models, to identify data points fitted poorly by the models and thereby suggesting other explanatory factors besides those incorporated into the models.

Tables 1 and 2 summarize the results of the first three types of statistical analysis. (Conclusions drawn from residuals are mentioned in the text below.) Signs of significant effects were mostly consistent between the analysis types. Effects on deforestation and on forest replacement had the same sign for six independent variables and opposite signs for one variable (latitude), whereas makatea and dust significantly affected forest replacement but not deforestation.

We now explain each independent variable whose possible role we had proposed at the outset, and the results for that variable. Two variables—rainfall, and latitude as a surrogate for temperature—were chosen because rainfall and temperature are primary determinants of plant growth rates. We reasoned that deforestation

should be less severe in areas where tree regrowth rates can keep pace with logging rates. A third variable, makatea, was chosen for reasons given below. Three variables— island age, volcanic ash fall-out and Asian dust fallout—were proposed to influence soil nutrient levels; regrowth is rapid on nutrient-rich soils (but those are also the soils preferred by farmers). The remaining three variables—elevation, area and distance—were proposed to have multiple indirect effects.

High rainfall was inversely associated with deforestation in all analyses (Table 1), and with replacement in one analysis (Table 2). That is, dry islands were more likely than wet islands to end up deforested. On 8 of the 12 islands for which we separately coded two locations, one location was much drier and also more deforested than the other, thereby offering a controlled natural experiment within an island. These effects are as expected; rainfall is often the most important single determinant of plant growth rates. In addition, low rainfall increases forest vulnerability to fire and hence to the formation of deforested grassland and fernland¹⁴.

Deforestation increased with latitude in all analyses, as expected from the decrease in temperature and hence in plant growth rates with latitude. In contrast, forest replacement decreased with latitude, undoubtedly because two of the most important introduced trees (breadfruit and Tahitian chestnut) are tropical species whose cultivation decreases with latitude.

Parts of 7 of our 69 islands consist of a terrain called makatea. This uplifted reef formation of sharp, fissured coral bears little soil and is painfully difficult to walk on. Not surprisingly, all seven islands provide controlled natural experiments: makatea terrain retained forests, whereas non-makatea terrain became deforested. Our statistical analysis therefore showed low forest replacement

Table 1 Significant predictors of deforestation

Independent variable	Bivariate regression		Multivariate regression			Tree	
	81	69	81, A	69	81, no A	81	69
Rainfall (log)	−1.9***	−1.6***	−1.3***	−1.3***	−1.03**	−, 1	−, 3
Latitude	+0.066***	+0.092***	+0.034**	+0.07***	+0.03*	+, 2	+, 1
Makatea							
Age	+0.53†	+0.53†		+0.29†			+, 2
Tephra	−0.46***	−0.37***	−0.61†		−0.79*	−, 3	
Dust							
Elevation (log)	−0.42*	−0.25†		−0.45**	−0.46**		
Area (log)	−0.31**	−0.24†	−0.26***				−, 2
Distance (log)			+0.22***	+0.18†	+0.20†		−, 3
Variance accounted for			0.62	0.75	0.59	0.65	0.72

The Methods section explains the nine independent variables and the bivariate regression, multivariate regression and multivariate tree analyses. Numbers in the two bivariate columns are bivariate regression coefficients; numbers in the multivariate columns are the corresponding multivariate regression coefficients. The signs in the tree columns are the sign of the relationship between deforestation and that independent variable; numbers in the tree columns denote the sequence in which predictor variables enter the tree (earlier-entering predictors, with lower numbers, are more important). Levels of statistical significance are as follows: ***, $P < 0.0001$; **, $P < 0.001$; *, $P < 0.01$; †, $P < 0.05$. Cells left blank failed to reach significance at $P < 0.05$. We used as alternatives a full data set of 81 entries and a reduced data set with 69 of those entries (numbers '81' and '69' in the first row). Multivariate regression of the full set with all nine independent variables (81, A) was repeated after dropping area as an independent variable (81, no A); that repetition was unnecessary for the 69-entry set because area proved not to be a significant predictor of deforestation in that set.

Table 2 Significant predictors of forest replacement

Independent variable	Bivariate regression		Multivariate regression				Tree	
	81	69	81, A	69, A	81, no A	69, no A	81	69
Rainfall (log)	−0.45†							
Latitude			−0.03**	−0.03*	−0.04***	−0.03†		
Makatea			−1.0†		−1.88***	−1.45**		
Age	+0.49*	+0.56***	−0.27*	+0.27†	−0.32*	−0.21‡		
Tephra	−0.58***	−0.34***	−0.72***	−0.48***	−0.94***	−0.74***	−, 2	−, 2
Dust	−0.003***	−0.002***	−0.002***	−0.001‡	−0.002***	−0.002**		
Elevation (log)	−0.84***	−0.45***	−0.58†	−0.28†	−1.2***	−0.96***		
Area (log)	−0.63***	−0.43***	−0.28*	−0.27*			−1, 2	−, 1
Distance (log)			+0.18*	+0.15†	+0.15†	+0.12§		
Variance accounted for			0.86	0.73	0.84	0.73	0.74	0.67

As Table 1, but for forest replacement instead of deforestation. Levels of statistical significance are as follows: ***, $P < 0.0001$; **, $P < 0.001$; *, $P < 0.01$; †, $P < 0.05$; ‡, $P = 0.06$; §, $P = 0.07$. Cells left blank failed to reach significance at $P < 0.05$.

associated with makatea, which is difficult to use for arboriculture. Our first three types of statistical analysis failed to show a significant association between makatea and deforestation, probably because the signal from just seven islands was too weak to emerge statistically in the whole data set. However, that association did appear in our statistical residuals: two of our islands with the highest percentage of makatea (Makatea Island and Niue) had the highest or nearly the highest negative residuals in both the multiple regression and tree analyses; that is, they were less deforested than expected just from effects of our other independent variables.

Island or terrain age is relevant because soil nutrients become lost from volcanic surfaces with time, especially by rain leaching^{15,16}. This effect, too, emerges from controlled natural experiments within islands: on Easter Island the oldest surface, the Poike Peninsula, became deforested several centuries before the younger remainder of the island⁷. In our statistical analyses deforestation increased with age, but the relation of forest replacement to age was inconsistent.

Those lost soil nutrients might become restored in three ways: volcanic ash fallout (aerial tephra), continental dust fallout and nesting seabird guano. The so-called Andesite line divides Pacific islands geographically into two groups: those whose volcanoes eject tephra carried by winds for up to 1,000 km and those whose volcanoes instead release lava with little aerial tephra. We did not realize this effect when we began our project, and it provided the biggest surprise: islands west of or near the Andesite line (that is, in or near the zone of aerial tephra) emerged with lower deforestation and forest replacement in all our types of non-residual statistical analyses.

Soil nutrients are also replenished by dust carried eastwards in the atmosphere from Central Asia^{15–17}. That dust fallout declines eastwards and southeastwards in the Pacific. Forest replacement decreased with increasing dust fallout. (We suspect a role of correlated geographical factors rather than of dust itself.) We could not detect an effect of dust on deforestation, but it might help to explain why our two islands farthest from Central Asia and hence with the lowest dust fallout, Easter and Mangareva, had among the largest positive residuals from our multiple regression model (that is, more deforested than predicted by our best-fit model).

Elevation was inversely associated with deforestation and forest replacement: high islands supported more forest and more native trees than low islands. This effect also emerges from natural experiments: virtually all high islands are forested with native trees at high elevation, even if they are deforested or support mainly introduced trees at low elevation. At least four factors are probably involved: orographic rain is generated at high elevations, descends in streams and thus makes the lowlands effectively wetter than indicated by lowland rainfall; nutrients and soil eroded at high elevation are similarly carried in streams to the lowlands; orographic rain captures atmospheric dust; and agriculture (hence land clearance) decreases with elevation because of cool temperatures (unfavourable for tropical crops), steep slopes and difficult access.

Area was inversely associated with deforestation and forest replacement: large islands retained more forest and more native tree species than small islands. Multiple factors probably contribute, including the fact that larger islands have greater habitat and tree species diversity (hence higher likelihood of some species being spared from logging), tracts of inaccessible land, and lower perimeter/area ratios (hence fewer coastal resources to support human population).

Distance (isolation from other islands) was positively associated in multiple regression analyses with deforestation and forest replacement, which tended to be most severe on remote islands. Reasons might include people on islands near other islands having the options of emigrating, trading or raiding instead of staying at home and having an impact on the forests, and low diversity in tree

species on remote islands decreasing the likelihood of any tree species being spared.

We can now reconsider why Easter Island⁴ suffered almost the most extreme deforestation and consequent social and population collapse of any Pacific island, even though the Polynesians who colonized Easter colonized hundreds of other islands without wreaking such extreme impacts. Our study suggests part of the answer to be Easter's extreme environmental fragility predisposing towards deforestation: of our 69 islands, it has the lowest tephra and dust fallout, the second greatest isolation and third highest latitude and no makatea, and is relatively low, small and dry. On the basis of those independent variables, our multiple regression and tree models predict correctly that Easter should have the third highest deforestation score, exceeded only by Necker and Nihoa, which also ended up completely deforested. That is, Easter's collapse was not because its people were especially improvident but because they faced one of the Pacific's most fragile environments.

Finally, we note seven avenues for refining our analysis. First, effects of island age, tephra fallout and dust fallout interact in a complex manner with rainfall: for example, nutrient leaching with age, and hence nutrient replenishment by tephra or dust, should be more significant on wetter islands¹⁶. Second, our measures of tephra, dust and age should be refined. Third, we lack a measure of nutrient inputs in seabird guano. Fourth, large-scale deforestation in the continental moist tropics can reduce convection and rainfall and thus cause positive feedback on the initial deforestation rate, but it is uncertain whether the same cycle operates on small islands. Fifth, a possible role of time since first colonization should be reconsidered; our preliminary analysis failed to detect a role. Sixth, although we have examined contributions of environmental differences to differences in deforestation and replacement, social differences surely also contribute, as Kirch discussed³. Indeed, our residual analyses suggest this conclusion: Tikopia and Tonga, whose societies Kirch noted as employing especially effective protective measures against deforestation, had two of the most negative residuals unexplained by our regression equations based solely on environmental factors, whereas Mangareva—cited by Kirch to illustrate ineffective social measures—had one of the highest positive residuals. Easter also had a high positive residual, which might reflect social pressures driving deforestation for transporting its famous stone statues. Last, our analysis could be extended to other societies, elsewhere in the world, noted either for deforestation or lack thereof; we think that we can discern straightforward extensions to the Anasazi, Japan, highland New Guinea, the Fertile Crescent and other cases^{1,6,8,9}. □

Methods

Values of the variables

The variables used are shown in Supplementary Table S1. Deforestation was scored on a five-point scale. A score of 1 represented no deforestation (no examples in our database; sole Pacific examples are some islands abandoned long ago or never settled by humans). A score of 2 meant densely forested; primary forests at higher elevation or on very steep slopes; mainly secondary forests with some introduced species at low elevations; no fire-associated vegetation. A score of 3 was as for 2 but with mainly introduced species in lowland secondary forests; much fire-associated grassland/fernland on ridges, slopes, and plateaus. A score of 4 indicated largely deforested; forests mainly on coastal plain, valley floors, and very steep slopes; fire-associated grassland/fernland covering all ridges and most slopes and plateaus. A score of 5 represented almost completely deforested; fire-associated grassland/fernland covering almost all areas not used as cropland.

Forest replacement was also scored on a five-point scale. A score of 1 indicated that introduced tree species comprised less than 10% of all tree individuals; 2, introduced tree species comprised 25–50% of forest tree individuals up to 600 m and less than 10% above 600 m; 3, introduced tree species comprised 50–75% of forest tree individuals up to 600 m and less than 10% above 600 m, with less land above 600 m than islands with a score of 2; 4, as 3 but introduced tree species comprised 75–100% of forest tree individuals up to 600 m; 5, there were few trees, and forest was mostly replaced by grasses and shrubs.

Sources for assessing deforestation and forest replacement were the earliest available first-hand descriptions by European visitors to each island (usually in the early-contact period), supplemented by paleoecological studies (pollen analysis).

We began with two alternative values for *A* (area) and *D* (distance): *A* of the individual island, or of all islands within 50 km; and *D* to nearest high island whose area is more than

25%, or more than 75%, of A of the target island. We then discarded the second measure in each case because of its high correlation with the first measure ($r = 0.94$ for the A measures; $r = 0.66$ for the D measures).

Lowland rainfall values were taken from atlases, journal articles, technical reports and other sources. Windward and leeward coasts were coded separately on 12 islands.

Our data table lists latitudes of island midpoints north or south of the Equator, but initial analyses showed no differences between effects of latitudes north or south; our subsequent analyses therefore used the absolute value of latitude without regard to whether north or south.

Makatea surface area as a percentage of total island surface area was estimated from geological maps.

Ages of volcanic rocks on islands were coded as follows: 1, young (less than 20,000 yr old); 2, intermediate (20,000–1,000,000 yr old); 3, old (more than 1,000,000 yr old); X, no exposed volcanic rock. These cutoffs were chosen on the basis of rates of nutrient leaching from volcanic rocks¹⁶. Islands formed from mosaics of terrains of different ages were assigned average values: for example, old islands with some young lava flows were scored $(3 + 1)/2 = 2$. Sources were journal articles based on K–Ar dating of volcanic rock.

Aerial tephra fallout was scored as follows: 1, low (islands more than 1,000 km east of the Andesite line); 2, moderate (islands east of the Andesite line but within 1,000 km of it); 3, high (islands west of the Andesite line).

Asian dust fallout values were taken or interpolated from refs 16 and 17 plus comments by J. Prospero.

Statistical analyses

Bivariate Spearman correlation coefficients, associated probability values (P) and the corresponding bivariate regression coefficients were calculated between the two outcome variables and nine independent variables.

Multivariate regression of each outcome variable on the independent variables employed robust linear regression methods¹² that minimized the absolute unsigned difference (not the squared difference) between observed and model-predicted values to yield least absolute deviation estimates of regression coefficients. We employed this method because the outcome variables are scored on a five-point scale and do not follow a gaussian distribution. Conventional multiple regression that instead minimizes the squared difference is in principle best for continuous measures and gaussian distributions; however, in practice robust and conventional multiple regression yielded virtually identical final models for our data sets (Supplementary Table S7). Initially, all independent variables were included in the model, and then variables with $P > 0.15$ were eliminated iteratively by backwards stepdown robust regression. Variables retained in the final models had $P \leq 0.05$, or $P < 0.0001$ in half of the cases. Log transforms were made on A, D, elevation and rainfall to obtain distributions more symmetrical and much closer to gaussian. These transformations also created a more linear relation to the outcomes.

Whereas multivariate regression assumes effects of independent variables to be linear, additive and non-interacting, the classification tree method does not make those assumptions and thus lends itself to detecting possible interactions¹³. Recursive binary partitioning is used to predict outcome scores by creating a tree consisting of groups of data points with outcome scores as different as possible between groups, and as similar as possible within groups. Each value of each independent variable is used to split the data into two groups provisionally, and that independent variable value is selected that creates the largest between-group difference relative to within-group variability. The process is repeated, splitting each such group into two further groups, until the final groups are 'pure' (all outcome values identical), sample sizes within a group are too small (less than five), and/or the mean difference between groups is not significant at the $P < 0.1$ level. In practice, all of our groups differed at least at $P < 0.03$, and usually at $P < 0.0001$. Variables entering the tree at earlier (higher) steps are more important than those entering at later (lower) steps.

Before reporting 'final' multiple regression and tree models for each data set (full or reduced) with or without A as an independent variable, we calculated model-predicted values of the two outcome variables at each data point, and computed and examined each residual value (predicted value minus actual value). Histograms and normal probability plots of these residual values showed that these errors had a bivariate symmetric unimodal distribution, demonstrating that the robust parametric approach was reasonable despite our semiquantitative deforestation and replacement scales.

For further details see Supplementary Methods.

Received 25 March; accepted 1 July 2004; doi:10.1038/nature02801.

1. Redman, C. L. *Human Impact on Ancient Environments* (Univ. Arizona Press, Tucson, 1999).
2. Kirch, P. V. *On the Road of the Winds* (Univ. California Press, Berkeley, 2000).
3. Kirch, P. V. Microcosmic histories: island perspectives on 'global' change. *Am. Anthropologists* **99**, 30–42 (1997).
4. Flenley, J. & Bahn, P. *The Enigmas of Easter Island* (Oxford Univ. Press, 2002).
5. Webster, D. *The Fall of the Ancient Maya* (Thames & Hudson, London, 2002).
6. Betancourt, J. L. & Van Devender, T. R. Holocene vegetation in Chaco Canyon, New Mexico. *Science* **214**, 656–658 (1981).
7. Mieth, A. & Bork, H.-R. Diminution and degradation of environmental resources by prehistoric land use on Poike Peninsula, Easter Island (Rapa Nui). *Rapa Nui J.* **17**, 34–41 (2003).
8. Totman, C. *The Green Archipelago: Forestry in Preindustrial Japan* (Univ. California, Berkeley, 1989).
9. Haberle, S. Dating the evidence for agricultural change in the Highlands of New Guinea: the last 2000 years. *Aust. Archaeol.* no. 47, 1–19 (1998).
10. Bellwood, P. *The Polynesians: Prehistory of an Island People* (Thames & Hudson, London, 1987).
11. Spriggs, M. *The Island Melanesians* (Blackwell, Oxford, 1997).
12. Jennrich, R. *Introduction to Computational Statistics: Regression Analysis*, Ch. 9 (Prentice Hall, Englewood Cliffs, New Jersey, 1995).

13. Brieman, L., Friedman, J., Olshen, R. & Stone, C. *Classification and Regression Trees* (Wadsworth, Belmont, California, 1984).
14. Barrau, J. *Subsistence Agriculture in Melanesia* (Bishop Mus. Bull. 219, Honolulu, 1958).
15. Austin, A. & Vitousek, P. Nutrient dynamics on a precipitation gradient in Hawai'i. *Oecologia* **113**, 519–529 (1998).
16. Chadwick, O. A., Derry, A. A., Vitousek, P. M., Huebert, B. J. & Hedin, L. O. Changing sources of nutrients during four million years of ecosystem development. *Nature* **397**, 491–497 (1999).
17. Ginoux, P. *et al.* Sources and distributions of dust aerosols simulated with the GOCART model. *J. Geophys. Res.* **106**, 20255–20273 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. Clague, C. Field, B. Huebert, B. Keating, P. Kirch, L. Kronke, A. Kurtz, E. Landaw, G. Merchant, J. Prospero and P. Vitousek for discussions, and J. Gornbein for the statistical design and analyses.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.R. (rolett@hawaii.edu).

Male mammals respond to a risk of sperm competition conveyed by odours of conspecific males

Javier delBarco-Trillo & Michael H. Ferkin

Department of Biology, University of Memphis, Ellington Hall, Memphis, Tennessee 38152, USA

Sperm competition occurs when a female copulates with two or more males and the sperm of those males compete within the female's reproductive tract to fertilize her eggs^{1,2}. The frequent occurrence of sperm competition has forced males of many species to develop different strategies to overcome the sperm of competing males^{1,3}. A prevalent strategy is for males to increase their sperm investment (total number of sperm allocated by a male to a particular female) after detecting a risk of sperm competition^{1,3,4}. It has been shown that the proportion of sperm that one male contributes to the sperm pool of a female is correlated with the proportion of offspring sired by that male^{5,6}. Therefore, by increasing his sperm investment a male may bias a potential sperm competition in his favour^{5,7,8}. Here we show that male meadow voles, *Microtus pennsylvanicus*, increase their sperm investment when they mate in the presence of another male's odours. Such an increase in sperm investment does not occur by augmenting the frequency of ejaculations, but by increasing the amount of sperm in a similar number of ejaculations.

Increases in sperm investment in response to a risk of sperm competition have been reported in many groups of animals, such as insects^{9,10}, fish¹¹ and birds¹². In mammals, similar reports are only indirect, in that risk of sperm competition was cued by the fact that males had not guarded the female before copulatory behaviour^{13,14}. Although mate guarding is an evolutionary response to minimize the risk of sperm competition, males ultimately assess particular risks of sperm competition in relation to other conspecific males⁵. Males may directly determine if another male has recently copulated with a particular female, for example, by detecting the presence of semen in her reproductive tract¹⁵. Often, however, a male may not know whether a female has copulated previously or whether a female will copulate with other males after he abandons her. In such cases, the most obvious way for a male to assess the risk of sperm

competition is by determining the presence of other males. This is because it is very likely that any promiscuous female, after copulating with a particular male, will also copulate with other nearby males¹⁶. Of the types of information that male mammals may use to determine the presence of other males and thus assess the risk of sperm competition, chemical signals (urine and other secretions used in chemical communication) are surely the most important, given that olfaction is the primary sensory modality among mammals^{17–19}. As far as we know, no previous studies have investigated whether male mammals use chemical signals to assess and respond to particular risks of sperm competition.

In the present two-part study, we determined whether male meadow voles use the odour cues of other males to assess risk of sperm competition and whether they respond by increasing their sperm investment. In the first part of the study, we did so by allowing ten male meadow voles to copulate with sexually receptive females in two different contexts (see Methods). In the 'risk of sperm competition' (RSC) context, the focal male was paired with a sexually receptive female in a cage that contained odours of another male. In the 'control' context, the focal male was paired with a female in a cage that did not contain odours of another male. Because meadow voles are highly promiscuous (litters in the wild are usually sired by more than one male²⁰), males regularly experience sperm competition. We also posit that because meadow voles can discriminate sex and reproductive status of conspecifics through their urine and other chemical signals¹⁹, they may be able to assess the presence of nearby males and the risk of sperm competition through their odours or scent marks. Thus, we proposed that male meadow voles would experience a higher risk of sperm competition in the RSC context than in the control context, thereby allocating more sperm in the RSC context than in the control context.

Each of the ten male meadow voles, when tested in the RSC context, investigated the odours from the donor male before (and also during) copulatory behaviour. This is a prerequisite to argue that male meadow voles increase their sperm investment in response to odours from conspecific males. More importantly, we found that male meadow voles increase significantly their sperm investment (Wilcoxon signed-rank test: $Z = -2.803$, $n = 10$, $P = 0.005$) when exposed to a risk of sperm competition conveyed by the odours of a conspecific male (Fig. 1). The average sperm investment in the RSC context was $168.99 \pm 18.12 \times 10^6$ spermatozoa (mean $\pm$ s.e.m.),

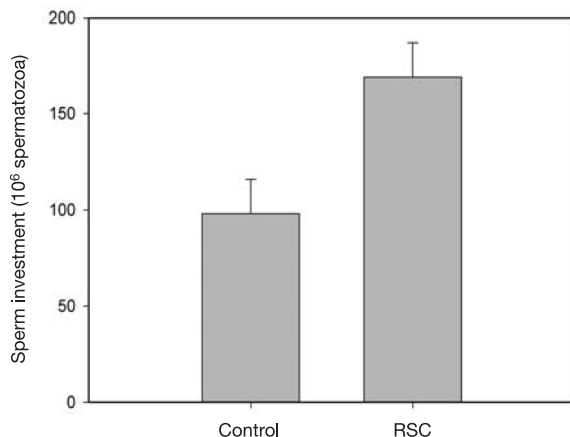


Figure 1 Sperm investment in the RSC and the control context. Sperm investment was significantly higher (Wilcoxon signed-rank test: $Z = -2.803$, $n = 10$, $P = 0.005$) when focal males were exposed to the odours of another male ($168.99 \pm 18.12 \times 10^6$ spermatozoa) than in the control context ($98.12 \pm 17.81 \times 10^6$ spermatozoa). Error bars indicate s.e.m.

whereas in the control context it was $98.12 \pm 17.81 \times 10^6$ spermatozoa.

Each of the ten male meadow voles increased their sperm investment in the RSC context in relation to the control context (Fig. 2a). The average increment was 116.32%. Although all males increased their sperm investment in the RSC context in relation to the control context, some males allocated more sperm in the control context than other males did in the RSC context, indicating a high variation in sperm investment among males in both the control and the RSC contexts (Fig. 2a). We determined that this high variation in sperm investment was due to differences in male body weight (Fig. 2b). A significant correlation between male body weight and sperm investment was found in the RSC context (Pearson's correlation: $r = 0.836$, $n = 10$, d.f. = 8, $P = 0.003$). In the control context a similar relationship was observed, although the correlation was not statistically significant ($r = 0.592$, $n = 10$, d.f. = 8, $P = 0.071$). When averaging the sperm investments of each male in both contexts, average sperm investment was significantly corre-

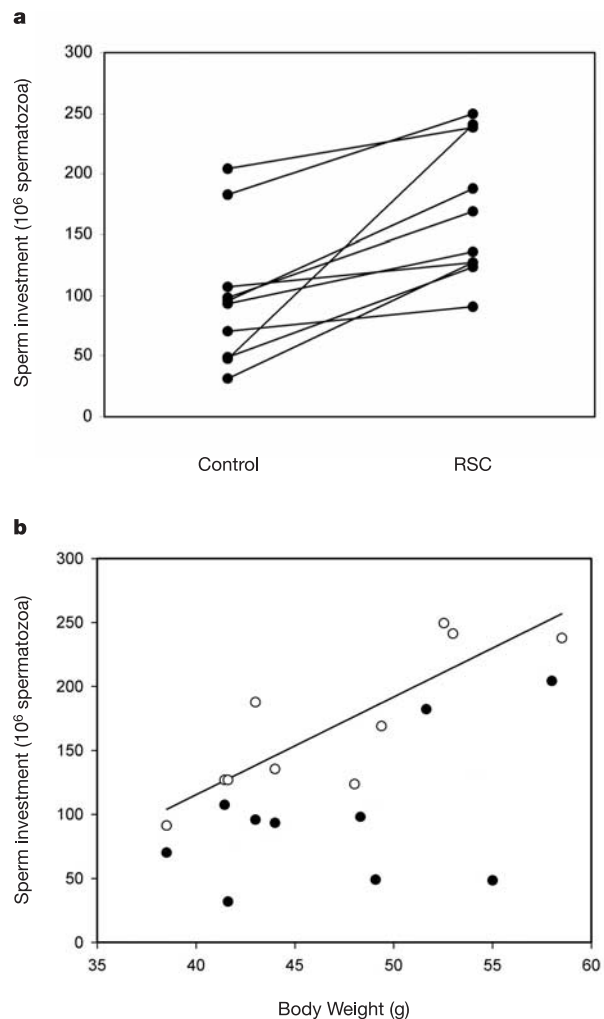


Figure 2 Sperm investment in relation to risk of sperm competition and male body weight. **a**, The sperm investments of each male in the control context and in the RSC context are connected by a line to show that all males increased their sperm investment in the RSC context in relation to the control context. The average increase was 116.32%. **b**, Sperm investment was positively correlated with body weight in the RSC context (open circles; Pearson's correlation: $r = 0.836$, $n = 10$, d.f. = 8, $P = 0.003$). In the control context a similar relationship can be observed, although it is not statistically significant (filled circles; $r = 0.592$, $n = 10$, d.f. = 8, $P = 0.071$).

lated to male body weight (Pearson's correlation: $r = 0.755$, $n = 10$, d.f. = 8, $P = 0.012$).

The second part of the study was a follow-up experiment that tested whether the increase in sperm investment observed in the RSC context may be induced by exposure to the odours of any heterospecific male or whether male meadow voles respond specifically to the chemical signals of conspecific males. To meet this objective, ten new male meadow voles were mated with receptive females in cages that contained the odours of a male mouse, *Mus musculus*. We then compared the sperm investment data of these new males with the data from the males used in the first part of the study. Male meadow voles allocated similar sperm investments (Mann–Whitney U test: $Z = -0.760$, $n = 10$, $P = 0.447$) in both the negative control (no odours of another male) and the positive control (male mouse odours; sperm investment, mean $\pm$ s.e.m. = $86.59 \pm 22.13 \times 10^6$ spermatozoa). Male meadow voles, however, allocated more sperm when exposed to the odours of a conspecific male (RSC context) than when exposed to the odours of a male mouse (Mann–Whitney U test: $Z = -2.508$, $n = 10$, $P = 0.012$). Therefore, we conclude that male meadow voles experience a risk of sperm competition, and respond to it by increasing their sperm investment when they come across the chemical signals of conspecific males, but not when they encounter the chemical signals of heterospecific males. Given that a particular area may be scent-marked by males from different species, it may be highly adaptive for males of a given species to specifically recognize and respond to the scent marks from conspecific males, because only conspecific males represent a sperm competition threat.

In species with a copulatory behaviour characterized by multiple ejaculations (that is, males ejaculating several times within the same female during the same copulatory bout)^{21,22}, such as the meadow vole²³, sperm investment may be adjusted by regulating either the number of sperm allocated in each ejaculation²⁴, or the total frequency of ejaculations^{5,22,25}. We found that male meadow voles do not increase their sperm investment by increasing their total frequency of ejaculations (Wilcoxon signed-rank test: $Z = -0.656$, $n = 10$, $P = 0.512$; Fig. 3). In both the control and the RSC contexts male meadow voles performed around six ejaculations (control context, mean $\pm$ s.e.m. = 5.63 ± 0.34 ; RSC context, 5.85 ± 0.21 ; range for both contexts, 4–7). Therefore, we conclude that in response to a risk of sperm competition, male meadow voles

increase the number of sperm allocated in some or all ejaculations without increasing the total frequency of ejaculations. This may be accomplished by increasing the contractility of the cauda epididymis and/or the vas deferens^{13,24}.

It has also been suggested that males may extend their 'total copulation interval' (amount of time performing copulatory behaviour) after detecting a risk of sperm competition. The argument is that the prolongation of copulatory behaviour may act as a form of mate guarding^{26,27}. However, our data do not support the mate-guarding hypothesis. Male meadow voles in the RSC context (mean $\pm$ s.e.m. = 116.19 ± 18.78 min) did not significantly prolong copulatory behaviour in relation to the control context (100.31 ± 16.05 min; Wilcoxon signed-rank test: $Z = -0.84$, $n = 10$, $P = 0.401$). Again, male meadow voles do not modify their copulatory behaviour when exposed to a risk of sperm competition. Instead, their only response is an increase in sperm investment.

This study highlights three important issues. First, male mammals may increase their sperm investment when they detect the presence of a conspecific male, in accordance with predictions from the sperm competition theory. Second, such an increase in sperm investment associated with a risk of sperm competition can be uncoupled from copulatory behaviour and attained merely through physiological changes. Finally, an association is presented between olfactory communication and sperm competition theory. Future research should focus on the type of information that males use to assess and respond to particular risks of sperm competition. □

Methods

Animals

All meadow voles used in the study were second- and third-generation offspring of field-caught animals, born and raised in a temperature-controlled room with a 14/10-h light/dark cycle. This photoperiod simulates a day length typical of the breeding season. Tests were always run during the first 2 h of the light cycle. All meadow voles were weaned at 19 days of age, housed with littermates until 34 days of age, and then housed singly in clear polycarbonate cages ($27 \times 16.5 \times 12.5$ cm). Cages contained hardwood shavings as bedding and cotton as nesting material. Food (Formulab Diet 5008, PMI Nutrition International Inc.) and tap water were provided *ad libitum*. Cage tops with filters were placed over each cage to avoid any colony odours from entering the cages. Also, all animals used in the study were visually isolated from each other by using wood partitions between cages. Thus, after they were 34 days of age the meadow voles used in our study were isolated from each other and the colony. The exception to this isolation occurred when the voles were 60–70 days of age and were mated with an opposite-sex conspecific. After the mating, the voles were returned to isolation until the start of the study, when they were approximately 120 days of age. This allowed the voles to have similar social and sexual experiences before the study. All animal procedures were approved by the IACUC of The University of Memphis.

Control and RSC contexts

Ten male and 20 female meadow voles were used for the pairings and ten other males as odour donors. The ten focal males were tested in both the control and the RSC contexts. To avoid any order effect, five males were randomly assigned first to the control context and 30 days later to the RSC context, whereas the other five males were assigned to the RSC context first and to the control context 30 days later. The 30-day period is enough time for small rodents to completely recover their normal epididymal sperm reserves²⁸. At the beginning of each trial, 20 g of experimental bedding were placed in the corner of a clean breeding cage ($37 \times 21 \times 15$ cm) lined with clean bedding. The experimental bedding was the only differing variable between the control and the RSC contexts. Twenty grams of bedding soaked in water were used in the control context. We could have used odours of the same focal male as an additional control; however, because male meadow voles scent mark quickly and abundantly when exposed to the stimuli of a receptive female²⁹, both controls are equivalent. In the RSC context, 20 g of soiled bedding (containing both urine and faeces) taken from the cage of a donor conspecific male were used. Soiled bedding was collected on the seventh day that the donor male had been in the cage. A different male donor was used in each trial, and none of the ten focal males were used as donors. Both control and RSC experimental beddings were always placed at the same position within the clean cage. A female meadow vole was introduced into the cage after the placement of the experimental bedding, and the focal male was introduced 5 min later. To increase female receptivity and ensure copulatory behaviour, all females were injected with 0.05 mg estradiol benzoate 3 days before pairing³⁰. Copulatory behaviour was taped and analysed afterwards to determine the ejaculation frequency and the total copulation interval²³. All males were allowed to reach sexual satiety, previously established for meadow voles to be 30 min without any intromission²³. Following male sexual satiety, the females were killed using an overdose of isoflurane. The female reproductive tract was removed, opened and all the semen diluted in 25 ml of distilled water. The solution was gently homogenized for

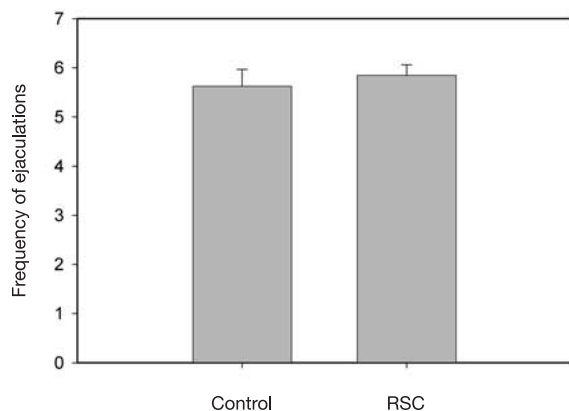


Figure 3 Frequency of ejaculations in the control and the RSC contexts. Frequency of ejaculations was not significantly different between the two contexts (Wilcoxon signed-rank test: $Z = -0.656$, $n = 10$, $P = 0.512$; range 4–7), being 5.63 ± 0.34 ejaculations in the control context and 5.85 ± 0.21 ejaculations in the RSC context. Male meadow voles respond to a risk of sperm competition by increasing their sperm investment (Fig. 1) without increasing their frequency of ejaculations. Error bars indicate s.e.m.

5 min, and four sperm counts were performed using a haemocytometer. The average of the four sperm counts was used to estimate the total number of sperm ejaculated by the male (or sperm investment).

Mouse bedding context

Ten male and ten female meadow voles were used in the second part of the study. The ten males were not used as focal or donor males in the first part of the study. Procedures were similar to both the control and RSC contexts, except for the use of 20 g of soiled bedding taken from the cage of a sexually active mouse. Donor mice were individually caged. A different mouse donor was used in each trial.

Received 7 June; accepted 13 July 2004; doi:10.1038/nature02845.

1. Birkhead, T. R. & Møller, A. P. (eds) *Sperm Competition and Sexual Selection* (Academic, San Diego, 1998).
2. Smith, R. L. (ed.) *Sperm Competition and the Evolution of Animal Mating Systems* (Academic, Orlando, 1984).
3. Wedell, N., Gage, M. J. G. & Parker, G. A. Sperm competition, male prudence and sperm-limited females. *Trends Ecol. Evol.* **17**, 313–320 (2002).
4. Parker, G. A., Ball, M. A., Stockley, P. & Gage, M. J. G. Sperm competition games: a prospective analysis of risk assessment. *Proc. R. Soc. Lond. B* **264**, 1793–1802 (1997).
5. Dewsbury, D. A. in *Sperm Competition and the Evolution of Animal Mating Systems* (ed. Smith, R. L.) 547–571 (Academic, New York, 1984).
6. Dziuk, P. J. Factors that influence the proportion of offspring sired by a male following heterospermic insemination. *Anim. Reprod. Sci.* **43**, 65–88 (1996).
7. Ginsberg, J. R. & Huck, U. W. Sperm competition in mammals. *Trends Ecol. Evol.* **4**, 74–79 (1989).
8. Schwagmeyer, P. L. & Foltz, D. W. Factors affecting the outcome of sperm competition in thirteen-lined ground squirrels. *Anim. Behav.* **39**, 156–162 (1990).
9. Wedell, N. & Cook, P. A. Butterflies tailor their ejaculate in response to sperm competition risk and intensity. *Proc. R. Soc. Lond. B* **266**, 1033–1039 (1999).
10. Gage, M. J. G. Risk of sperm competition directly affects ejaculate size in the Mediterranean fruit fly. *Anim. Behav.* **42**, 1036–1037 (1991).
11. Pilastro, A., Scaggiante, M. & Rasotto, M. B. Individual adjustment of sperm expenditure accords with sperm competition theory. *Proc. Natl Acad. Sci. USA* **99**, 9913–9915 (2002).
12. Pizzari, T., Cornwallis, C. K., Lovlie, H., Jakobsson, S. & Birkhead, T. R. Sophisticated sperm allocation in male fowl. *Nature* **426**, 70–74 (2003).
13. Baker, R. R. & Bellis, M. A. *Human Sperm Competition* 203–227 (Chapman & Hall, New York, 1995).
14. Bellis, M. A., Baker, R. R. & Gage, M. J. G. Variation in rat ejaculates consistent with the Kamikaze-Sperm hypothesis. *J. Mamm.* **71**, 479–480 (1990).
15. Lezama, V., Orihuela, A. & Angulo, R. Sexual behavior and semen characteristics of rams exposed to their own semen or semen from a different ram on the vulva of the ewe. *Appl. Anim. Behav. Sci.* **75**, 55–60 (2001).
16. Dewsbury, D. A. Ejaculate cost and male choice. *Am. Nat.* **119**, 601–610 (1982).
17. Johnston, R. E. in *Pheromones and Reproduction in Mammals* (ed. Vandenbergh, J. G.) 3–37 (Academic, New York, 1983).
18. Brown, R. E. & Macdonald, D. W. (eds) *Social Odours in Mammals* 1–36 (Oxford Univ. Press, Oxford, 1985).
19. Ferkin, M. H. & Johnston, R. E. Meadow voles, *Microtus pennsylvanicus*, use multiple sources of scent for sex recognition. *Anim. Behav.* **49**, 37–44 (1995).
20. Boonstra, R., Xia, X. & Pavone, L. Mating system of the meadow vole, *Microtus pennsylvanicus*. *Behav. Ecol.* **4**, 83–89 (1993).
21. Dewsbury, D. A. Patterns of copulatory behavior in male mammals. *Q. Rev. Biol.* **47**, 1–33 (1972).
22. Dewsbury, D. A. Diversity and adaptation in rodent copulatory behavior. *Science* **190**, 947–954 (1975).
23. Gray, G. D. & Dewsbury, D. A. A quantitative description of the copulation behaviour of meadow voles (*Microtus pennsylvanicus*). *Anim. Behav.* **23**, 261–267 (1975).
24. Pound, N. Effects of morphine on electrically evoked contractions of the vas deferens in two congeneric rodent species differing in sperm competition intensity. *Proc. R. Soc. Lond. B* **266**, 1755–1758 (1999).
25. Parker, G. A. Sperm competition games: raffles and roles. *Proc. R. Soc. Lond. B* **242**, 120–126 (1990).
26. Birkhead, T. R. & Møller, A. P. *Sperm Competition in Birds: Evolutionary Causes and Consequences* (Academic, London, 1992).
27. Simmons, L. W. *Sperm Competition and its Evolutionary Consequences in the Insects* 144–187 (Princeton Univ. Press, Princeton, 2001).
28. Robb, G. W., Amann, R. P. & Killian, G. J. Daily sperm production and epididymal sperm reserves of pubertal and adult rats. *J. Reprod. Fert.* **54**, 103–107 (1978).
29. Ferkin, M. H., Lee, D. N. & Leonard, S. T. The reproductive state of female voles affects their scent marking behavior and the responses of male conspecifics to such marks. *Ethology* **110**, 257–272 (2004).
30. Dewsbury, D. A. & Baumgardner, D. J. Studies of sperm competition in two species of muroid rodents. *Behav. Ecol. Sociobiol.* **9**, 121–133 (1981).

Acknowledgements This work was supported by a Sigma Xi Grant-in-Aid of Research to J.d.-T., a NIH Grant to M.H.F. and a NIH Grant to the Tennessee Mouse Genome Consortium. J.d.-T. designed and carried out the experiments, analysed the data and wrote the paper; M.H.F. assisted in writing of the final drafts of the paper, as well as providing research facilities, animals and material.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.d.-T. (jtrillo@memphis.edu).

Theory predicts the uneven distribution of genetic diversity within species

Erik M. Rauch^{1,2} & Yaneer Bar-Yam²

¹MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar Street, Cambridge, Massachusetts 02139, USA

²New England Complex Systems Institute, 24 Mount Auburn Street, Cambridge, Massachusetts 02138, USA

Global efforts to conserve species have been strongly influenced by the heterogeneous distribution of species diversity across the Earth. This is manifest in conservation efforts focused on diversity hotspots^{1–3}. The conservation of genetic diversity within an individual species^{4,5} is an important factor in its survival in the face of environmental changes and disease^{6,7}. Here we show that diversity within species is also distributed unevenly. Using simple genealogical models, we show that genetic distinctiveness has a scale-free power law distribution. This property implies that a disproportionate fraction of the diversity is concentrated in small sub-populations, even when the population is well-mixed. Small groups are of such importance to overall population diversity that even without extrinsic perturbations, there are large fluctuations in diversity owing to extinctions of these small groups. We also show that diversity can be geographically non-uniform—potentially including sharp boundaries between distantly related organisms—without extrinsic causes such as barriers to gene flow or past migration events. We obtained these results by studying the fundamental scaling properties of genealogical trees. Our theoretical results agree with field data from global samples of *Pseudomonas* bacteria. Contrary to previous studies⁸, our results imply that diversity loss owing to severe extinction events is high, and focusing conservation efforts on highly distinctive groups can save much of the diversity.

Our approach is to use simulations and analytic studies of the genealogical tree of a population, a method known as coalescent theory^{9–12}. We test the robustness of our results in part by varying the model details but primarily by obtaining analytic results (see Box 1). For studies of well-mixed populations, we use the Wright–Fisher model^{13,14}: each organism is the offspring of a randomly chosen parent of the previous generation. For spatial populations (Fig. 1), the parent of an organism is either the previous organism at that site, or one of the neighbours on a square lattice with equal probability for all possible parents. Each of the parent–child links is an opportunity for genetic mutation. Thus, the expected genetic distance between two individuals is proportional to the time to their common ancestor (their relatedness), and the expected total diversity is proportional to the number of links traced back from the current population to the common ancestor. (If the rate of mutation is sufficiently high that multiple mutations can occur at a single locus within a genealogical tree, a correction is necessary: the total diversity is $D(B) \approx (\mu/\mu_L) (1 - \exp(-\mu_L B))$, where B is the total branch length, μ is the per-genome mutation rate and μ_L is the probability of a particular mutation at a particular locus.) The total diversity has been characterized for well-mixed populations¹⁵. Here we study the way that the diversity is distributed within populations.

We first consider the genetic uniqueness of an individual or group; that is, the degree to which an individual or group is distinctive from all other individuals in the population, for example, the minimal time to an ancestor common with any other individual. The average of this quantity for well-mixed populations has been calculated^{14,16}, but its distribution has not been studied. As can be

seen from the genealogy, some of the genetic diversity in the population is found only in particular individuals; other diversity is common to smaller or larger groups that share a common ancestor. We find that uniqueness of individuals has a scale-free (power law) distribution (Fig. 2a). This property is highly robust and is valid for both well-mixed and spatial populations (exponents -2.93 and -2.85 (dashed line), respectively); however, spatial populations contain more highly distinctive individuals and groups, as the tail of the distribution extends farther.

This distribution also applies to the uniqueness of groups composed of genetically related individuals (see Box 1). Sexual reproduction increases the likelihood of moderate values of uniqueness in relation to high values. However, the power law tail remains, so that highly distinctive individuals exist even in a sexually reproducing population (Fig. 2b). The distribution for well-mixed sexual populations is similar to that of spatial sexual populations, although the shorter tail enhances the effect of averaging. Incorporation of spatially uniform selection into the model has little effect unless there are mutations that are advantageous enough to sweep the population in a short time (periodic selection or genetic hitch-hiking). This results in a cutoff of the most ancient part of the tree at the time of the most recent strongly advantageous mutation, thus cutting off the tail of the uniqueness distribution. Otherwise, spatially uniform fitness differences have minimal impact. Selection that varies in space (or multiple local niches) would increase the tail of the uniqueness distribution by providing more highly unique individuals associated with long genealogies in different regions. The power law distribution of diversity suggests that much of the genetic diversity is found in a small portion of the population.

We compared these results with genetic data obtained previously¹⁷ on field samples of *Pseudomonas* bacteria. From the dendrogram of 250 samples, taken from different locations around the world, we obtained the distribution of uniqueness of the

samples, as described in the Supplementary Information, and shown as circles in Fig. 3. The results are long-tailed and can be fitted by a power law distribution. To compare more precisely the genetic data with our theoretical calculations, however, we must include the effect of sampling. Sampling makes the distribution longer-tailed, corresponding to a greater proportion of individuals that are more unique with respect to the sampled population. We directly simulated the ancestral tree of the samples, initializing the simulation by representing organisms at the specific geographical locations where the *Pseudomonas* samples were obtained. The result is shown as a dashed line in Fig. 3. At each step of the simulation, moving backwards in time, a lineage performs a random walk on a 50×25 lattice, staying in place or moving to a neighbouring site. At its destination, with a certain probability, p_c , it coalesces with other lineages at that site. Details are given in the Supplementary Information. p_c , the number of simulation time steps N_T corresponding to one unit T_0 of biological time, and the lattice size are adjustable parameters. The parameters were set by a simple fitting procedure that adjusts the intercepts of $U(T)$ at the U and T axes and accounts for the small effect of sampling on the low- T values of the curve.

The distribution of uniqueness in well-mixed and spatial populations leads to significant temporal variation. It has been noted that the time to the common ancestor can have intermittent large jumps¹⁸. Here we study the size distribution of fluctuations in diversity and its relationship to the uniqueness distribution. In Fig. 2c we show the diversity, measured by the total branch length of the genealogical tree, over time. The fluctuations are large: diversity grows gradually owing to the accumulation of mutations, but often decreases markedly in a single generation; in the figure, diversity decreases in one generation by up to about two-thirds of the total. The sudden decreases are due to the extinction of lines of descent that have accumulated many unique mutations. The distribution of the losses of diversity (Fig. 2a) has a power law tail with the same exponent as the uniqueness distribution, consistent with the loss of uniqueness of randomly chosen individuals. The flattening of the distribution for small losses reflects averaging due to the loss of multiple lines of descent in a particular generation. Other diversity measures such as nucleotide diversity¹⁹ show similarly large fluctuations.

Box 1

Analytic result

The distribution of genetic uniqueness, and hence also of fluctuations in the total branch length of the genealogy, can be understood as follows. The probability $P(U > u)$ that an individual has uniqueness greater than u is the probability that its lineage, traced backwards, never exists on a site that has another lineage for all time $T \leq u$. In the well-mixed model, the probability that no other lineage jumps to a particular site is $((N-1)/N)^{p(T)N} \approx \exp(-p(T))$, where $p(T)$ is the fraction of individuals living at time T in the past that have descendants in the present. $p(T)$ is approximately $\frac{a}{T}$, with a measured from simulations of $p(T)$ to be 1.95, and expected analytically to be 2 for the well-mixed case¹⁴. This gives:

$$P(U > u) \approx \prod_{T=1}^u \exp\left(-\frac{a}{T}\right) = \exp\left(-a \sum_{T=1}^u \frac{1}{T}\right) \\ \sim \exp(-a \log(bu)) \sim u^{-a}$$

Thus the exponent depends on the coefficient a of $p(T)$. The probability density is $P(u) = -dP(U > u)/du$, giving

$$P(u) \sim u^{-a-1} \sim u^{-2.95}$$

consistent with both the well-mixed and spatial simulations of $P(u)$.

The scale-free distribution of uniqueness also applies to subgroups defined by a given level of relatedness T_g . For each individual, define the subgroup it belongs to by the identity of its ancestor T_g generations ago. We define the uniqueness u_g of the group to be the uniqueness of the ancestor. The genealogical tree of the ancestors of these groups have the same properties as that of the present population of individuals, only starting with a lower value of $p(T)$. Thus, their uniqueness follows the same power law distribution. For the same reason, the distribution is not affected by the level of resolution in genetic distances (the smallest measurable difference).

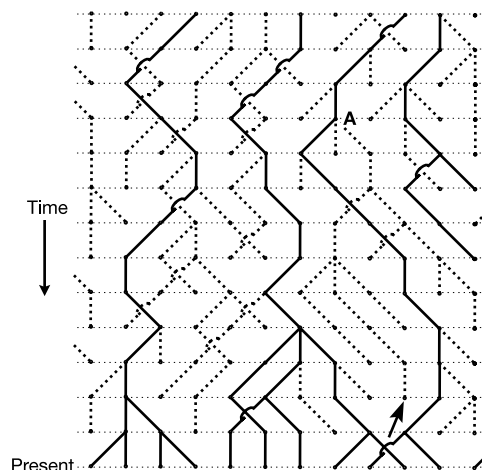


Figure 1 Section of a genealogical tree for a one-dimensional population (part of a much larger tree). Time proceeds down the page, with each horizontal row representing a (non-overlapping) generation and offspring connected to their parent by a line. The tree of the currently living individuals is shown as solid lines; the ancestry of those that have no descendants in the present, and thus do not contribute to diversity, is shown as dashed lines. At the arrow, a lineage goes extinct, causing the loss of accumulated differences on the line of descent from A, the most recent ancestor with descendants in the present.

We can also characterize how diversity is distributed spatially. In Fig. 1, it can be seen that organisms are often closely related to their neighbours in a spatial population, but that neighbours can also be quite distantly related. Previous work has only considered one-dimensional populations²⁰. Using a simulation in two dimensions, Fig. 2d shows that there are specific patch-like areas, which represent groups separated by a large genealogical distance from the rest of the population. The shaded areas are separated by 10,000

generations from the rest of the population; the grey areas are separated from the black ones by 2,000 generations. Continued simulation shows that specific boundaries move and disappear, but distinctive features exist at any particular time. In this simulation, dispersal of offspring is limited to neighbouring sites. If offspring disperse farther away but still near to their parents, types become interspersed locally and the patches do not have sharp boundaries. Still, as long as the dispersal distance is small compared with the size of the habitat, there remains a spatial patchy structure of genealogically well-separated types. If recombination occurs, there are still such patterns but they are distinct for each independently inherited part of the genome.

Explanations for the occurrences of divergent populations in particular spatial areas, and boundaries between types, are often sought in habitat variation, barriers that prevent organism motion or in a past migration event^{21–23}. The model shows that divergent patches can arise from the internal structure of the genealogical tree. Divergent populations are not necessarily confined to a single area; alleles can be geographically widespread in a population even if it is not well-mixed. Thus the spatial patterns of genetic variation in homogeneous habitats must be considered before making inferences about the properties and history of a population. Regarding global microbial populations in particular, the model suggesting that these populations are unaffected by geographic barriers or limited dispersal, and vary spatially only due to selection²⁴, has been challenged by recent work^{17,25}. Here we consider this question only for a specific bacterial species; for others our work does not exclude the cosmopolitan limit of globally mixed populations through which only selection produces local variation. However, our work implies that observations that seem to suggest a globally mixed population, or that variation arises from selection acting on types with otherwise cosmopolitan distributions, must be considered in light of the intrinsic diversity distribution of genealogies. Moreover, even if global dispersal dominates on longer timescales and eliminates biogeographic variation observable in studies with

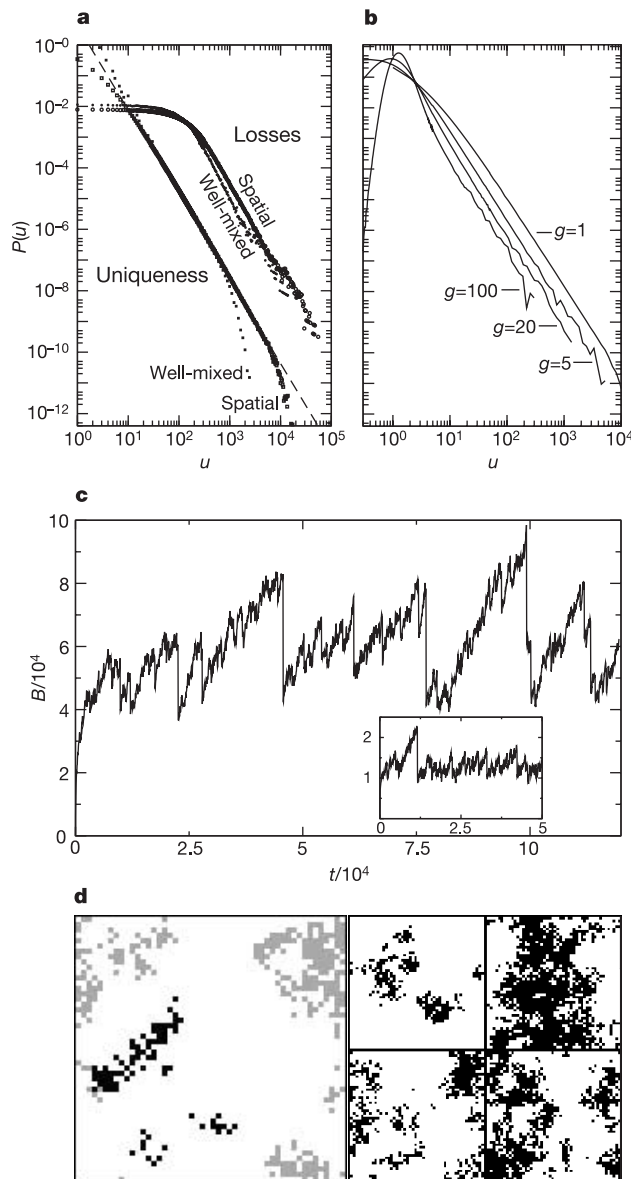


Figure 2 Distribution of diversity and its fluctuation. **a**, Distribution $P(u)$ of genetic uniqueness of individuals, and of losses of diversity in a single generation in well-mixed (small symbols) and spatial populations (larger symbols). Horizontal axis represents uniqueness in generation of divergence. **b**, Distribution of uniqueness in a spatial sexual population for different numbers of independently inherited segments of the genome $g = 1, 5, 20$ and 100 . The total genome size is fixed. Linkage would reduce the difference from the $g = 1$ case. **c**, Time series of the diversity of a spatial population. Inset shows a similar well-mixed case. **d**, Simulated pattern of the most divergent groups (grey and black, see text) in a spatial habitat (lattice size $l = 50$). The four panels on the right show the most divergent group in the later evolution of the same population (black), at intervals of 2,000 generations.

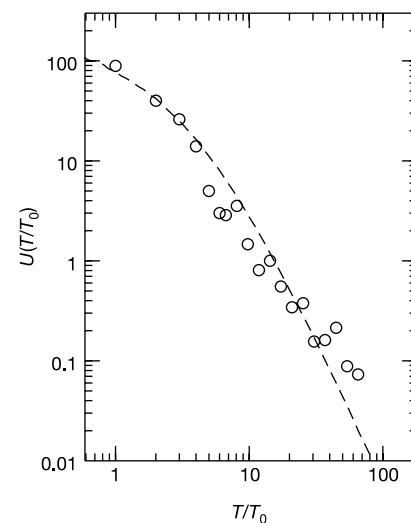


Figure 3 Comparison of theoretical values of uniqueness with data from field populations. Circles represent $U(T/T_0)$, the number of samples with a uniqueness of T/T_0 , for a sampled *Pseudomonas* population. The dashed line is an average over 1,000 spatial simulations. We normalize T by dividing by T_0 , the time to the smallest genetic difference considered. Sampling causes a shallower slope than for the whole population (Fig. 2), and this slope is matched by the simulation. In the simulation, T_0 corresponds to 160 time steps, the lattice size is 50×25 , and the coalescence probability, p_c , is 0.15. Details are given in the Supplementary Information.

coarse genetic resolution, the finer-scale differences now becoming observable by genetic fingerprinting are more likely to display biogeography. Extrinsic factors, such as local selection or habitat boundaries, can be distinguished from intrinsic genealogical diversity. For sexually reproducing populations, for extrinsically imposed diversity, spatial boundaries of independently inherited parts of the genome would tend to coincide, whereas for intrinsic genealogical diversity they would tend to be different. For microbial populations, where non-systematic recombination cannot provide the same information, short generation times may allow space-time sampling to study the biogeographical dynamics.

Although the model used here is simple, our analysis shows that the results are robust, and independent of the level of genetic resolution. In summary, our results show that genetic diversity is very unevenly distributed. A small fraction of a population is responsible for a disproportionate fraction of the diversity. Diversity has its own internal dynamics distinct from external influences such as habitat change and species interactions. Increases happen only gradually, but large decreases may occur without an extrinsic perturbation.

It was recently suggested⁸, on the basis of the analysis of model phylogenetic trees (similar to our genealogical trees), that extinction of 95% of species would leave 80% of the tree of life (total diversity) retained, and that as a result ecological planning to preserve diversity is not constructive. These results arise because random losses, even when high, are unlikely to remove all individuals belonging to a deep branch of the tree even when it forms a small proportion of the population, thus preserving most of the diversity.

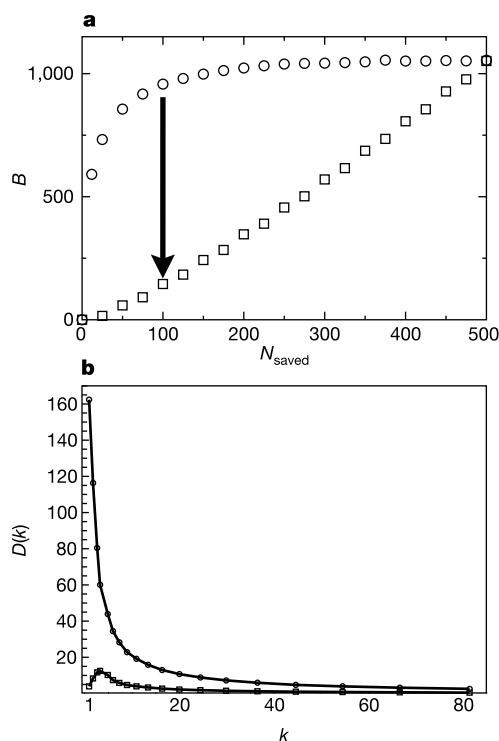


Figure 4 Diversity retained after an extinction episode in a well-mixed population. **a**, The average diversity (total branch length B) remaining after the population is reduced from 500 to the value N_{saved} (horizontal axis). Circles show the values immediately after the extinction and squares the value subsequently reached. The arrow indicates the diversity lost after the initial extinction for a loss of 80% of the population. **b**, $D(k)$, the number of mutations carried by exactly k individuals (multiplicity or redundancy). The area under this curve is the total diversity. $D(k)$ is shown just after the extinction event (upper curve), and in the long term.

In contrast, our studies suggest that conservation planning is important and can enable substantial improvement of diversity preservation. We note that for spatial populations, the loss of some of the habitat leads to a greater probability of eliminating all members of a divergent group because members of divergent groups are spatially clustered, therefore potentially leading to a much greater loss of diversity. Even for well-mixed populations, however, the small immediate loss (Fig. 4a; circles, similar to Fig. 2b of ref. 8) is followed by a much greater loss over time (Fig. 4a; squares) owing to the vulnerability of residual divergent groups to extinction. Simulations show that most of the subsequent loss occurs within 20 generations. The vulnerability of residual small groups of related organisms can be seen from a plot revealing the multiplicity (redundancy²⁶), k , of mutations (Fig. 4b). For the case shown, almost 25% of the total diversity just after the extinction is found in only one or two individuals, and most of the diversity lost is found in less than ten individuals. This shows that ensuring the reproduction of rare types by conservation planning^{5,6,27,28} during or even just after an extinction episode can markedly improve diversity retention. □

Received 26 April; accepted 11 June 2004; doi:10.1038/nature02745.

1. Prendergast, J. R., Quinn, R. M., Lawton, J. H., Eversham, B. C. & Gibbons, D. W. Rare species, the coincidence of diversity hotspots and conservation strategies. *Nature* **365**, 335–337 (1993).
2. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
3. Gaston, K. J. Global patterns in biodiversity. *Nature* **405**, 220–227 (2000).
4. Ehrlich, P. R. & Wilson, E. O. Biodiversity studies: science and policy. *Science* **253**, 758–761 (1991).
5. Faith, D. P. Genetic diversity and taxonomic priorities for conservation. *Biol. Conserv.* **68**, 69–74 (1994).
6. Amos, A. & Balmford, A. When does conservation genetics matter? *Heredity* **87**, 257–265 (2001).
7. Frankham, R. *et al.* Do population size bottlenecks reduce evolutionary potential? *Anim. Conserv.* **2**, 255–260 (1999).
8. Nee, S. & May, R. M. Extinction and the loss of evolutionary history. *Science* **278**, 692–694 (1997).
9. Hudson, R. R. Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* **7**, 1–44 (1990).
10. Barton, N. H. & Wilson, I. Genealogies and geography. *Phil. Trans. R. Soc. Lond. B* **349**, 49–59 (1995).
11. Notohara, M. The structured coalescent process with weak migration. *J. Appl. Prob.* **38**, 1–17 (2001).
12. Wilkins, J. F. & Wakeley, J. The coalescent in a continuous, finite, linear population. *Genetics* **161**, 873–888 (2002).
13. Wright, S. Isolation by distance. *Genetics* **28**, 114–138 (1943).
14. Fisher, R. A. The distribution of gene ratios for rare mutations. *Proc. R. Soc. Edinb.* **50**, 205–220 (1930).
15. Watterson, G. A. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276 (1975).
16. Wakeley, J. & Takahashi, T. Gene genealogies when the sample size exceeds the effective size of the population. *Mol. Biol. Evol.* **20**, 208–213 (2003).
17. Cho, J. C. & Tiedje, J. M. Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. *Appl. Environ. Microbiol.* **66**, 5448–5456 (2000).
18. Watterson, G. A. Mutant substitutions at linked nucleotide sites. *Adv. Appl. Prob.* **14**, 206–224 (1982).
19. Nei, M. *Molecular Evolutionary Genetics* (Columbia University Press, New York, 1987).
20. Irwin, D. E. Phylogeographic breaks without geographic barriers to gene flow. *Evolution* **56**, 2383–2394 (2002).
21. Hewitt, G. M. The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913 (2000).
22. Willis, K. J. & Whittaker, R. J. The refugial debate. *Science* **287**, 1406–1407 (2000).
23. Cann, R. Genetic clues to dispersal in human populations: retracing the past from the present. *Science* **291**, 1742–1748 (2001).
24. Finlay, B. J. Global dispersal of free-living microbial eukaryote species. *Science* **296**, 1061–1063 (2002).
25. Whitaker, R. J., Grogan, D. W. & Taylor, J. W. Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301**, 976–978 (2003).
26. Bar-Yam, Y. Multi-scale variety in complex systems. *Complexity* **9**, 37–45 (2004).
27. Crozier, R. H. Preserving the information content of species: genetic diversity, phylogeny, and conservation worth. *Annu. Rev. Ecol. Syst.* **28**, 243–268 (1997).
28. Moritz, C. Defining evolutionarily significant units for conservation. *Trends Ecol. Evol.* **9**, 373–375 (1994).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was funded in part by the National Science Foundation. We are indebted to S. Hubbell, S. Pimm, J. Wakeley, M. Kardar and C. Goodnight for comments. We thank J.-C. Cho and J. Tiedje for comments and for providing the original figure with data from ref. 17.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.R. (rauch@necsi.org).

A frequency-dependent switch from inhibition to excitation in a hippocampal unitary circuit

Masahiro Mori, Mathias H. Abegg, Beat H. Gähwiler & Urs Gerber

Brain Research Institute, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

The hippocampus, a brain structure essential for memory and cognition, is classically represented as a trisynaptic excitatory circuit. Recent findings challenge this view, particularly with regard to the mossy fibre input to CA3, the second synapse in the trisynaptic pathway¹. Thus, the powerful mossy fibre input to CA3 pyramidal cells might mediate both synaptic excitation and inhibition^{2,3}. Here we show, by recording from connected cell pairs in rat entorhinal–hippocampal slice cultures, that single action potentials in a dentate granule cell evoke a net inhibitory signal in a pyramidal cell. The hyperpolarization is due to disynaptic feedforward inhibition, which overwhelms monosynaptic excitation. Interestingly, this net inhibitory synaptic response changes to an excitatory signal when the frequency of presynaptic action potentials increases. The process responsible for this switch involves the facilitation of monosynaptic excitatory transmission coupled with rapid depression of inhibitory circuits. This ability to immediately switch the polarity of synaptic responses constitutes a novel synaptic mechanism, which might be crucial to the state-dependent processing of information in associative hippocampal networks.

We studied, in hippocampal slice cultures, the unitary response in a CA3 pyramidal cell (PC) evoked by action potentials in a connected granule cell (GC). Connectivity of GCs in our slice cultures approaches that reported *in vivo*, with no evidence for sprouting^{4,5} (Fig. 1a, Supplementary Fig. 1a). In six GCs we measured a mossy fibre length of $2,008 \pm 219 \mu\text{m}$, bearing 11.8 ± 1.3 large mossy fibre terminals with a diameter of $5.9 \pm 0.3 \mu\text{m}$, of which 70.4% gave rise to 1.6 ± 0.1 filopodial extensions in CA3 and with 13.6 ± 2.2 large mossy fibre terminals giving rise to 1.7 ± 0.1 filopodial extensions in the hilus. These values compare well with those reported for rat *in vivo*, in which mossy fibre length was $3,236 \mu\text{m}$, the number of large mossy fibre terminals was 12.3 (range 10–18) with a diameter ranging between 4 and $10 \mu\text{m}$ with 2.3 filopodial extensions in CA3, and with 7–12 large mossy fibre terminals in the hilus².

A single action potential in the GC induced a biphasic response in the PC consisting of a brief excitatory postsynaptic current (EPSC) reliably followed by a pronounced inhibitory postsynaptic current (IPSC) (12 of 13; Fig. 1b, Supplementary Fig. 2). The resulting charge transfer was $1,134.1 \pm 378.7 \text{ fC}$ outwards ($n = 12$), indicative of a net inhibitory synaptic response. This inhibitory response might arise through disynaptic inhibition^{2,6–8} or from the simultaneous release of glutamate and GABA from mossy fibre terminals³. To distinguish between these possibilities we blocked excitatory glutamatergic signalling with the α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA)/kainate receptor antagonist 1,2,3,4-tetrahydro-6-nitro-2,3-dioxo-benzo[f]-quinoxaline-7-sulphonamide (NBQX; $10 \mu\text{M}$), which abolished both the excitatory and the inhibitory components of the response ($n = 12$; Fig. 1b). Thus, the inhibitory component reflects disynaptic feedforward or feedback inhibition. Because the unitary charge transfer in a PC evoked by a CA3 interneuron was $327.1 \pm 65.3 \text{ fC}$ ($n = 8$; five interneurons in CA3 stratum lucidum, three interneurons in stratum oriens), we estimate that at least four inter-

neurons discharged to evoke the inhibitory response in a PC after a single action potential in a GC.

Evidence for a monosynaptic excitatory connection is provided by the low variability in latency (jitter: $0.53 \pm 0.10 \text{ ms}$, $n = 7$) between the GC action potential and the postsynaptic response (Figs 1b, c and 2a) and the one-to-one transmission at high frequencies (Figs 2e, f and 3). Furthermore, evoking pharmacologically isolated *N*-methyl-D-aspartate (NMDA) receptor-mediated EPSPs, which are below firing threshold, resulted in short latency responses in 6 of 16 GC–PC pairs, which is consistent with a monosynaptic connection (Fig. 1c). The lack of a response in ten pairs reflects the absence of NMDA receptors at a significant proportion of mossy fibre synapses^{9,10}. Monosynaptic AMPA/kainate receptor-mediated EPSCs from GC–PC pairs had a response latency of $6.7 \pm 0.7 \text{ ms}$, a mean peak amplitude of $-163.0 \pm$

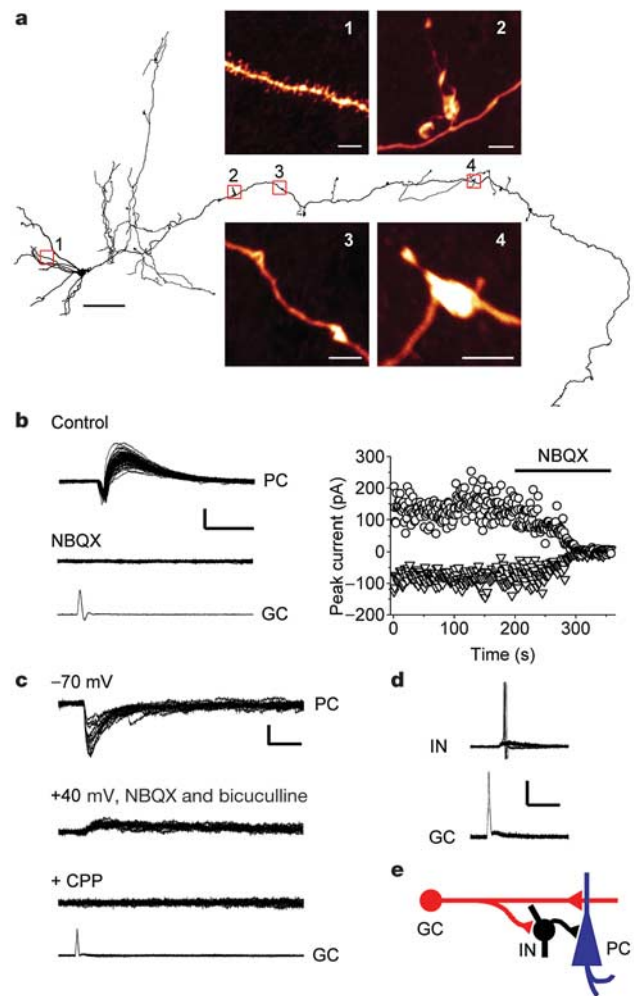


Figure 1 Unitary responses induced by single GC action potentials. **a**, A GC in slice culture (see Supplementary Fig. 1a). Boxed regions show a spiny GC dendrite (1), large mossy fibre boutons with filopodial extensions (2 and 4), and *en passant* varicosities (3). **b**, Left, single action potentials at 1 Hz in a GC (bottom) evoked biphasic currents in a PC at -70 mV (top) (see Supplementary Fig. 2), sensitive to $10 \mu\text{M}$ NBQX (middle). Right, amplitudes of EPSCs (triangles) and IPSCs (circles) plotted against time. **c**, Under suppression of IPSCs in a recorded PC (see Methods), a single action potential (1 Hz, bottom) evokes synaptic currents mediated by AMPA/kainate receptors (top), and NMDA receptors (after $20 \mu\text{M}$ NBQX and $40 \mu\text{M}$ bicuculline, upper middle), sensitive to $20 \mu\text{M}$ CPPene (lower middle). **d**, GC-evoked synaptic potentials and action potentials in an interneuron (IN) (-65 mV) (see Supplementary Fig. 1b–i). **e**, Unitary circuit for mossy fibre transmission. Vertical bars: 100 pA (PC), 100 mV (GC) (**b**); 50 pA (PC), 100 mV (GC) (**c**); 50 mV (**d**). Horizontal bars: left $100 \mu\text{m}$, images $5 \mu\text{m}$ (**a**); 20 ms (**b–d**).

23.3 pA at -70 mV, a 20–80% rise time of 1.6 ± 0.2 ms and a decay time constant of 12.8 ± 1.1 ms ($n = 7$; Fig. 2a). The reliability of monosynaptic EPSC transmission depended strongly on the frequency with which single presynaptic action potentials were elicited (Fig. 2b) but was not modified by depolarizing the presynaptic GC by 10 mV ($83.0 \pm 6.1\%$ and $82.8 \pm 6.7\%$ for failure rates with a 3-s interval at presynaptic potentials of -70 mV and -60 mV, respectively; $P = 0.98$, $n = 4$). This result indicates that conduction failures¹¹ do not contribute significantly to the observed failures in synaptic transmission, although this mechanism cannot be ruled out. The failure rate of the first EPSP in a train decreased with repetitive trains of action potentials, presumably because of persisting residual calcium. Paired-pulse facilitation in PCs was apparent at an interspike interval shorter than 100 ms (2.54 ± 0.48 at 50 ms, $n = 5$; Fig. 2c, d). When trains of action potentials were induced in a GC, EPSC facilitation changed to a gradual depression with increasing spike frequency (Fig. 2e, f).

Activation of a single GC evoked subthreshold responses in targeted PCs, but had to have induced action potentials in the CA3 interneurons comprising the feedforward circuit (Fig. 1e). Indeed, a single action potential in a GC elicited an action potential in targeted postsynaptic interneurons ($30.3 \pm 6.1\%$ of trials at -60 mV, $n = 4$, three of five stratum lucidum interneurons; one of three stratum oriens interneurons; Fig. 1d, Supplementary Fig. 1b–i) as reported previously for hippocampal interneurons^{12,13}. Monosynaptic EPSCs from GC–interneuron pairs had a response latency of 7.0 ± 0.7 ms, a mean peak amplitude of -56.9 ± 17.7 pA

at -70 mV, a 20–80% rise time of 1.1 ± 0.3 ms, and a decay time constant of 7.3 ± 1.4 ms ($n = 5$).

GCs normally fire at low frequencies (less than 1 Hz), which can, however, increase greatly (10–40 Hz) when rats enter the place field of a GC¹⁴. To examine frequency-dependent effects on transmission, we recorded responses in a PC to trains of 15 action potentials induced in a synaptically coupled GC at frequencies ranging from 10 to 40 Hz. At 10 Hz the synaptic response to each presynaptic action potential consisted of a biphasic EPSP/IPSP sequence (Fig. 3a), with inhibition predominating as determined by calculating the area of each response (Fig. 3b). Facilitation of both EPSPs and IPSPs was observed. At firing frequencies of 20 Hz and greater, the initial facilitation of EPSPs and IPSPs was succeeded by a rapid and complete depression of the IPSP component, resulting in monophasic EPSPs (Fig. 3a–d). The progressive depression of the IPSPs was associated with a corresponding increase in duration of the EPSPs. The mean half-width of the EPSPs by the 13th to 15th action potential was 3.7 ± 0.2 ms at 10 Hz, compared with 6.8 ± 0.4 ms at 40 Hz ($n = 7$, $P < 0.05$). This frequency-dependent shift in IPSP-dominant to EPSP-dominant responses could be repeatedly reproduced by allowing a 30-s interval between trains and was unaffected by application of the GABA_B receptor antagonist CGP62349 ($3 \mu\text{M}$; $P = 0.98$, $n = 5$) or the metabotropic glutamate receptor antagonist α -methyl-4-carboxyphenylglycine (MCPG; 0.5 mM, $P = 0.96$, $n = 6$). Strong extracellular stimulation leads to chloride redistribution, reducing the driving force for inhibition¹⁵. However, when we induced trains of action potentials in a single mossy fibre, there was no apparent shift in the reversal potential for inhibition (control, -84.3 ± 1.7 mV; after 15 action potentials at 40 Hz, -85.0 ± 1.8 mV; $P = 0.77$, $n = 4$), indicating

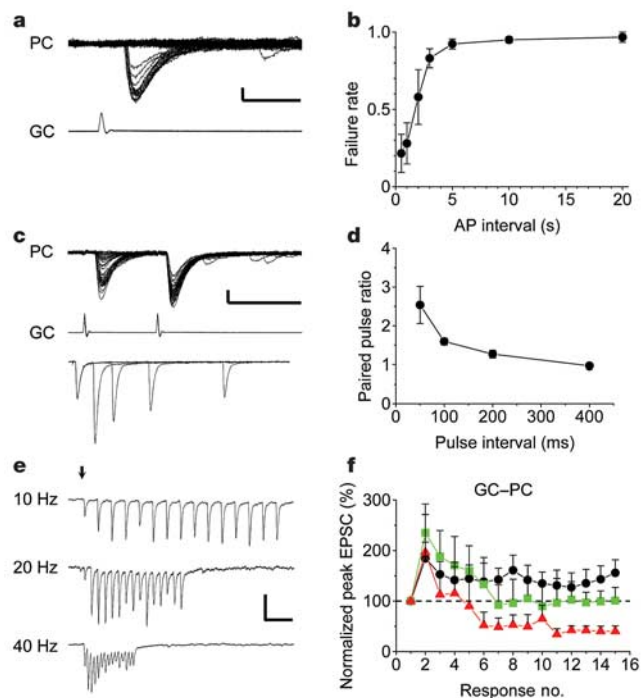


Figure 2 Frequency-dependent properties of monosynaptic mossy fibre transmission to PCs. **a**, Monosynaptic currents evoked by single GC action potentials at 0.5 Hz in a PC with inhibition blocked (-70 mV). **b**, Transmission failure decreases as action potential frequency increases ($n = 5$). **c**, Paired-pulse facilitation at an interspike interval of 50 ms at 0.5 Hz in the same cell pair as in **a**. Averaged paired-pulse responses at 50–400 ms normalized to the first response amplitude (bottom). **d**, Plot of paired-pulse ratio against pulse interval ($n = 5$). **e**, Monosynaptic currents in a PC during trains of 15 presynaptic action potentials with increasing frequencies. Traces are averages (three to five sweeps). **f**, Peak amplitude of EPSCs at 10 Hz (black), 20 Hz (green) and 40 Hz (red) plotted against evoked response number (1–15) ($n = 6$). Vertical bars: 50 pA (PC), 100 mV (GC) (**a**); 100 pA (PC), 100 mV (GC) (**c**); 100 pA (**e**). Horizontal bars: 20 ms (**a**); 50 ms (PC and GC), 200 ms (bottom) (**c**); 200 ms (**e**). Error bars, s.e.m.

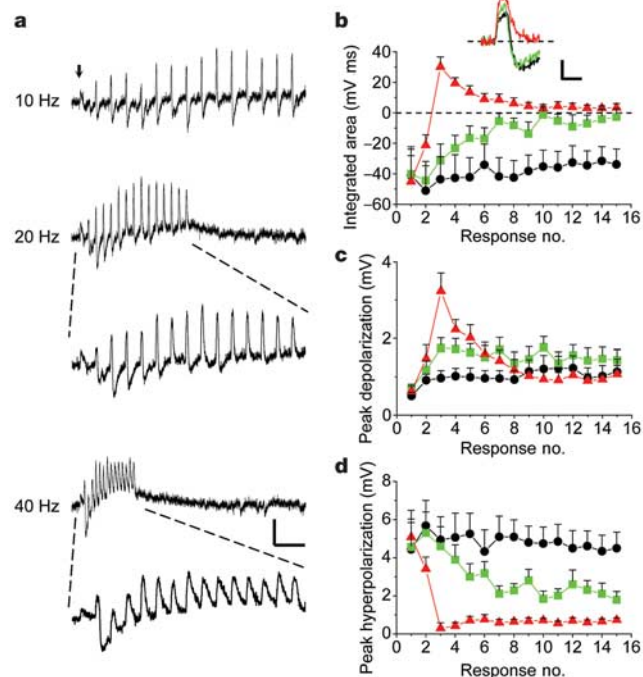


Figure 3 Switch in unitary postsynaptic potentials from inhibitory-dominant to excitatory-dominant with increasing frequency. **a**, Synaptic potentials in a PC (-70 mV) evoked by 15-action potential trains with increasing frequencies. Averages of four or five sweeps are shown. **b**, Integration of postsynaptic potentials for 20 ms after each action potential at 10 Hz (black), 20 Hz (green) and 40 Hz (red) plotted against response number (1–15) ($n = 7$). Inset, superimposition of fourth responses (**a**). **c**, **d**, Peak depolarization (**c**) and hyperpolarization (**d**) at 10 Hz (black), 20 Hz (green) and 40 Hz (red) plotted against response number ($n = 7$). Vertical bars, 3 mV (**a**); 2 mV (**b**). Horizontal bars, 200 ms; expanded traces 100 ms (20 Hz) and 50 ms (40 Hz) in **a**, 10 ms in **b**. Error bars, s.e.m.

no significant redistribution of chloride. To characterize the mechanism underlying the rapid depression of disynaptic IPSPs at high firing rates, we examined the frequency response of the two synapses involved in the feedforward inhibition of PCs (Fig. 1e). For the monosynaptic EPSCs mediating spike transmission (Fig. 1d) at the GC–interneuron synapse ($n = 5$; interneurons in stratum lucidum), EPSCs progressively facilitated at 10–20 Hz, whereas at 40 Hz the responses were depressed after the eighth action potential (Fig. 4a, b). A similar analysis at the interneuron–PC synapse ($n = 5$, interneurons in stratum lucidum) showed that GABAergic transmission was depressed rapidly at all frequencies tested (Fig. 4c, d). Thus both the GC–interneuron and interneuron–PC synapses contribute to the rapid depression of disynaptic IPSPs in response to high-frequency discharge of GCs (Fig. 4e).

These results indicate first that unitary mossy fibre transmission to PCs induces not only a monosynaptic excitatory response but also a powerful disynaptic inhibitory response resulting in net inhibition, and second that an increase in the frequency of GC firing results in a switch from an inhibitory-dominant to an excitatory-dominant postsynaptic response.

The observation that extracellular stimulation of the dentate gyrus evokes a biphasic EPSP/IPSP response in PCs¹⁶ and the demonstration of an extensive innervation of interneurons by the mossy fibres² has indicated an important role for feedforward inhibition in the modulation of activity in the CA3 area^{6–8}. Our recordings from functionally connected GC–PC pairs confirmed this hypothesis and allowed us to determine that in rat hippocampal slice cultures four or more interneurons typically contribute to this feedforward pathway. In contrast, the monosynaptic excitatory pathway generally consists of one giant synapse on a PC, containing

30–40 active zones^{2,17}. We calculated a unitary conductance of the monosynaptic glutamatergic mossy fibre response of 2.3 ± 0.3 nS ($n = 7$; Fig. 2a). This value translates to a release of 18 vesicles in response to a single presynaptic action potential, assuming a quantal conductance at room temperature of 0.13 nS (ref. 10). Under our recording conditions (34 °C), quantal size will be greater, indicating that a single action potential is likely to release at most one vesicle per active zone. At high frequencies of presynaptic action potential firing, Ca^{2+} influx enhanced by presynaptic action potential broadening¹⁸ and Ca^{2+} release from intracellular stores^{19,20} lead to a marked increase in Ca^{2+} in the terminal, which will release a large fraction of the vesicles from the readily releasable pool. However, because of the significant size of this pool (1,400) (ref. 21) it is unlikely that vesicular depletion is induced by short trains of action potentials. Thus, depression of the monosynaptic mossy fibre EPSCs (Fig. 2e, f) is probably due to other mechanisms such as postsynaptic AMPA/kainate receptor desensitization.

A unitary local neuronal network composed of one GC, a targeted PC and several interneurons will serve as a high-pass filter for incoming temporal events in which high-frequency signals will be transferred as a rate code, whereas less frequent basal events will be processed into a temporally precise code. Evidence for enhanced excitatory signals onto PCs in response to high-frequency discharge was observed in a study *in vivo* in which trains of presynaptic spikes discharged PCs more effectively²². In our study, stimulation of a single GC never evoked a suprathreshold EPSP in the PC at holding potentials between -70 and -60 mV. However, when PCs receive excitatory drive from the local associative network^{23,24}, the membrane potential will be closer to threshold, allowing unitary mossy fibre EPSPs to trigger action potentials. The frequency-dependent switch from inhibition to excitation might be important in certain forms of modulation of synaptic strength such as long-term depression and long-term potentiation^{25,26}. Thus, frequency-dependent modulation of synaptic transmission seems to be an important principle in the regulation of brain circuits^{27,28}. □

Methods

Preparation of slice cultures

Slice cultures were prepared from 6-day-old Wistar rat pups killed by decapitation in accordance with a protocol approved by the Veterinary Department of the Canton of Zurich, and maintained as described previously²⁹. In brief, 400- μm -thick hippocampal slices including the entorhinal cortex were attached to glass coverslips with the use of clotted chicken plasma, placed in sealed test tubes with serum-containing medium, and kept in a roller-drum incubator at 36 °C for 21–28 days.

Electrophysiological recordings

Cultures were transferred to a recording chamber mounted on an upright microscope (Axioskop FS1; Zeiss) and superfused with an external solution containing 148.8 mM Na^+ , 2.7 mM K^+ , 149.2 mM Cl^- , 2.8 mM Ca^{2+} , 2.0 mM Mg^{2+} , 11.6 mM HCO_3^- , 0.4 mM H_2PO_4^- , 5.6 mM D-glucose and 10 mg l^{-1} phenol red (pH 7.4). All experiments were performed at 34 °C. Recordings were obtained from PCs, interneurons in area CA3a and b, and GCs with patch pipettes (2–5 M Ω) using an Axopatch 200B amplifier (Axon Instruments). For current-clamp recording, pipettes were filled with 135 mM potassium gluconate, 5 mM KCl, 10 mM Hepes, 1 mM EGTA, 2 mM MgATP, 5 mM creatine phosphate (CrP), 0.4 mM GTP, 0.07 mM CaCl_2 (pH 7.2). For voltage-clamp recording, patch pipettes were filled with 121.6 mM CsF, 8.4 mM CsCl, 10 mM HEPES, 10 mM EGTA, 1 mM picrotoxin, 2 mM MgATP, 5 mM CrP, 0.4 mM GTP (pH 7.2), except that the potassium solution as above was used for the voltage-clamp recordings shown in Figs 1b and 4c. To suppress IPSCs in recorded PCs, we used a solution with fluoride as the major intracellular anion containing the GABA_A receptor antagonist picrotoxin and adjusted the equilibrium potential for chloride to correspond to the holding potential of -70 mV. Membrane potentials were corrected for liquid junction potentials. Series resistance (5–15 M Ω) was compensated but by not more than 40–90% to avoid unstable recordings, and cells were excluded if a change of more than 20% occurred. Presynaptic action potentials were evoked by injecting depolarizing current (1 ms, 1.5–2.0 nA) at 0.5 Hz unless otherwise mentioned. By recording first from a postsynaptic PC and then systematically scanning the dentate gyrus with a potassium puff electrode (140 mM K^+) to induce localized activation ('potassium puff search'), the success rate of obtaining a monosynaptic connection between a GC and a PC was increased up to 10%.

Imaging and image analysis

For staining cells, 0.1–0.2% biocytin or tetramethylrhodamine (Mioruby; Molecular Probes) was added to the pipette solutions. After experiments, slices were fixed in 4%

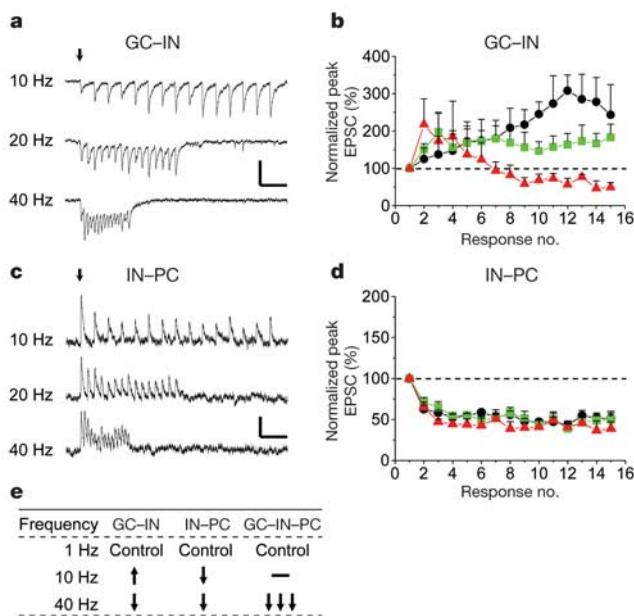


Figure 4 Depression of both GC–interneuron (GC–IN) and interneuron–PC (IN–PC) synapses at high frequency contributes to the switch from inhibitory to excitatory CA3 responses. **a**, Modulation of interneuron EPSCs with increasing action potential frequencies (-70 mV). **b**, EPSC amplitude normalized to the first EPSC at 10 Hz (black), 20 Hz (green) and 40 Hz (red) plotted against response number (1–15) in GC–interneuron pairs ($n = 5$). **c**, IPSCs at interneuron–PC synapses were depressed at 10–40 Hz. **d**, IPSC amplitude normalized to the first IPSC at 10 Hz (black), 20 Hz (green) and 40 Hz (red) plotted against response number in interneuron–PC pairs ($n = 5$). **e**, Summary of frequency-dependent modulation for GC–interneuron, interneuron–PC and GC–interneuron–PC synapses. Traces are averages of three to five sweeps. Vertical bars, 50 pA (**a**); 20 pA (**c**). Horizontal bars, 200 ms. Error bars, s.e.m.

buffered paraformaldehyde, removed from the coverslip, permeabilized for 6–12 h, incubated for 15 min in Alexa Fluor 488 streptavidin (Molecular Probes), washed again and mounted. Eight-bit image stacks were acquired on a confocal microscope (SP2; Leica) using the 40× HCXP APO (numerical aperture 1.25) oil-immersion objective. Voxel size in the xyz dimension was $0.4 \times 0.4 \times 0.5 \mu\text{m}^3$ ($0.05 \times 0.05 \times 0.5 \mu\text{m}^3$ for high resolution). Measures of axonal arborization and diameter of the large mossy terminals were performed on montages of maximal intensity projections (Supplementary Fig. 1a) using ImageJ software (National Institute of Mental Health, Bethesda, Maryland). No further filtering or image processing was used.

Drugs and chemicals

MCPG and NBQX were purchased from Tocris Cookson, and ATP, CrP, EGTA, GTP, (–)-bicuculline methochloride and 3,3′-diaminobenzidine from Sigma/Fluka. (E)-4-(3-phosphonoprop-2-enyl)piperazine-2-carboxylic acid (CPPene) and CGP62349 were kindly provided by Novartis AG.

Data acquisition and analysis

Signals were filtered at 2 or 5 kHz, digitally recorded on a computer by using CLAMPEX 7 software (Axon Instruments) and stored on tape for later analysis. The amplitude, latency and kinetics were determined as described elsewhere³⁰. To quantify the synaptic responses evoked by each action potential during a train, the peak amplitude and the area of the response were measured from the baseline directly preceding each EPSP. The standard deviation of the latencies was used to calculate the jitter. For calculation of the paired-pulse ratios, averages (including failures) were obtained. Numerical data in the text are expressed as means $\pm$ s.e.m.

Received 9 January; accepted 12 July 2004; doi:10.1038/nature02854.

1. Bischofberger, J. & Jonas, P. Two B or not two B: differential transmission at glutamatergic mossy fiber–interneuron synapses in the hippocampus. *Trends Neurosci.* **25**, 600–603 (2002).
2. Ácsády, L., Kamondi, A., Sik, A., Freund, T. & Buzsáki, G. GABAergic cells are the major postsynaptic targets of mossy fibers in the rat hippocampus. *J. Neurosci.* **18**, 3386–3403 (1998).
3. Walker, M. C., Ruiz, A. & Kullmann, D. M. Monosynaptic GABAergic signaling from dentate to CA3 with a pharmacological and physiological profile typical of mossy fiber synapses. *Neuron* **29**, 703–715 (2001).
4. Zimmer, J. & Gähwiler, B. H. Cellular and connective organization of slice cultures of the rat hippocampus and fascia dentata. *J. Comp. Neurol.* **228**, 432–446 (1984).
5. Frotscher, M. & Gähwiler, B. H. Synaptic organization of intracellularly stained CA3 pyramidal neurons in slice cultures of rat hippocampus. *Neuroscience* **24**, 541–551 (1988).
6. Frotscher, M. Mossy fiber synapses on glutamate decarboxylase-immunoreactive neurons: evidence for feed-forward inhibition in the CA3 region of the hippocampus. *Exp. Brain Res.* **75**, 441–445 (1989).
7. Schwartzkroin, P. A., Scharfman, H. E. & Sloviter, R. S. Similarities in circuitry between Ammon's horn and dentate gyrus: local interactions and parallel processing. *Prog. Brain Res.* **83**, 269–286 (1990).
8. Spruston, N., Lübke, J. & Frotscher, M. Interneurons in the stratum lucidum of the rat hippocampus: an anatomical and electrophysiological characterization. *J. Comp. Neurol.* **385**, 427–440 (1997).
9. Siegel, S. J. et al. Regional, cellular, and ultrastructural distribution of N-methyl-D-aspartate receptor subunit 1 in monkey hippocampus. *Proc. Natl Acad. Sci. USA* **91**, 564–568 (1994).
10. Jonas, P., Major, G. & Sakmann, B. Quantal components of unitary EPSCs at the mossy fibre synapse on CA3 pyramidal cells of rat hippocampus. *J. Physiol. (Lond.)* **472**, 615–663 (1993).
11. Debanne, D., Guéneau, N. C., Gähwiler, B. H. & Thompson, S. M. Action-potential propagation gated by an axonal I_A -like K^+ conductance in hippocampus. *Nature* **389**, 286–289 (1997).
12. Miles, R. Synaptic excitation of inhibitory cells by single CA3 hippocampal pyramidal cells of the guinea-pig *in vitro*. *J. Physiol. (Lond.)* **428**, 61–77 (1990).
13. Debanne, D., Guéneau, N. C., Gähwiler, B. H. & Thompson, S. M. Physiology and pharmacology of unitary synaptic connections between pairs of cells in areas CA3 and CA1 of rat hippocampal slice cultures. *J. Neurophysiol.* **73**, 1282–1294 (1995).
14. Jung, M. W. & McNaughton, B. L. Spatial selectivity of unit activity in the hippocampal granular layer. *Hippocampus* **3**, 165–182 (1993).
15. Thompson, S. M. & Gähwiler, B. H. Activity-dependent disinhibition. I. Repetitive stimulation reduces IPSP driving force and conductance in the hippocampus *in vitro*. *J. Neurophysiol.* **61**, 501–511 (1989).
16. Brown, T. H. & Johnston, D. Voltage-clamp analysis of mossy fiber synaptic input to hippocampal neurons. *J. Neurophysiol.* **50**, 487–507 (1983).
17. Chicurel, M. E. & Harris, K. M. Three-dimensional analysis of the structure and composition of CA3 branched dendritic spines and their synaptic relationships with mossy fiber boutons in the rat hippocampus. *J. Comp. Neurol.* **325**, 169–182 (1992).
18. Geiger, J. R. & Jonas, P. Dynamic control of presynaptic Ca^{2+} inflow by fast-inactivating K^+ channels in hippocampal mossy fiber boutons. *Neuron* **28**, 927–939 (2000).
19. Llano, I. et al. Presynaptic calcium stores underlie large-amplitude miniature IPSCs and spontaneous calcium transients. *Nature Neurosci.* **3**, 1256–1265 (2000).
20. Liang, Y., Yuan, L. L., Johnston, D. & Gray, R. Calcium signaling at single mossy fiber presynaptic terminals in the rat hippocampus. *J. Neurophysiol.* **87**, 1132–1137 (2002).
21. Hallermann, S., Pawlu, C., Jonas, P. & Heckmann, M. A large pool of releasable vesicles in a cortical glutamatergic synapse. *Proc. Natl Acad. Sci. USA* **100**, 8975–8980 (2003).
22. Henze, D. A., Wittner, L. & Buzsáki, G. Single granule cells reliably discharge targets in the hippocampal CA3 network *in vivo*. *Nature Neurosci.* **5**, 790–795 (2002).
23. Miles, R. & Wong, R. K. Excitatory synaptic interactions between CA3 neurones in the guinea-pig hippocampus. *J. Physiol. (Lond.)* **373**, 397–418 (1986).
24. Nakazawa, K. et al. Requirement for hippocampal CA3 NMDA receptors in associative memory recall. *Science* **297**, 211–218 (2002).
25. Yeckel, M. F., Kapur, A. & Johnston, D. Multiple forms of LTP in hippocampal CA3 neurons use a common postsynaptic mechanism. *Nature Neurosci.* **2**, 625–633 (1999).
26. Kobayashi, K. & Poo, M. M. Spike train timing-dependent associative modification of hippocampal CA3 recurrent synapses by mossy fibers. *Neuron* **41**, 445–454 (2004).

27. Galarreta, M. & Hestrin, S. Frequency-dependent synaptic depression and the balance of excitation and inhibition in the neocortex. *Nature Neurosci.* **1**, 587–594 (1998).
28. Varela, J. A., Song, S., Turrigiano, G. G. & Nelson, S. B. Differential depression at excitatory and inhibitory synapses in visual cortex. *J. Neurosci.* **19**, 4293–4304 (1999).
29. Gähwiler, B. H., Thompson, S. M., McKinney, R. A., Debanne, D. & Robertson, R. T. in *Culturing Nerve Cells* 2nd edition (eds Banker, G. & Goslin, K.) 461–498 (MIT Press, Cambridge, Massachusetts, 1998).
30. Feldmeyer, D., Egger, V., Lübke, J. & Sakmann, B. Reliable synaptic connections between pairs of excitatory layer 4 neurones within a single 'barrel' of developing rat somatosensory cortex. *J. Physiol. (Lond.)* **521**, 169–190 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank H. Blum, S. Giger, H. Kasper, L. Rietschin, and R. Schöb for technical assistance, and P. Streit for help in image processing. This work was funded by the Swiss National Science Foundation, the NCCR on Neural Plasticity and Repair, and the Hartmann Müller Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.M. (mori@hifo.unizh.ch).

Regulation of B-cell survival by BAFF-dependent PKC δ -mediated nuclear signalling

Ingrid Mecklenbräuer¹, Susan L. Kalled², Michael Leitges³, Fabienne Mackay⁴ & Alexander Tarakhovskiy¹

¹Laboratory of Lymphocyte Signaling, The Rockefeller University, New York, New York 10021, USA

²Biogen Idec, Inc., Cambridge, Massachusetts 02142, USA

³Max Planck Institute for Experimental Endocrinology, D 30625 Hannover, Germany

⁴Garvan Institute of Medical Research, St Vincent Hospital, Darlinghurst, NSW 2010, Sydney, Australia

Approximately 65% of B cells generated in human bone marrow are potentially harmful autoreactive B cells¹. Most of these cells are clonally deleted in the bone marrow, while those autoreactive B cells that escape to the periphery are anergized or perish before becoming mature B cells^{2–5}. Escape of self-reactive B cells from tolerance permits production of pathogenic auto-antibodies⁶; recent studies suggest that extended B lymphocyte survival is a cause of autoimmune disease in mice and humans⁷. Here we report a mechanism for the regulation of peripheral B-cell survival by serine/threonine protein kinase C δ (PKC δ): spontaneous death of resting B cells is regulated by nuclear localization of PKC δ that contributes to phosphorylation of histone H2B at serine 14 (S14-H2B). We show that treatment of B cells with the potent B-cell survival factor BAFF ('B-cell-activating factor belonging to the TNF family') prevents nuclear accumulation of PKC δ . Our data suggest the existence of a previously unknown BAFF-induced and PKC δ -mediated nuclear signalling pathway which regulates B-cell survival.

The increased lifespan of B cells in transgenic mice that over-express the B-cell survival factor BAFF is associated with development of lupus-like autoimmunity^{8–10}. We recently described a lupus-like autoimmune disease with defective B-cell tolerance in mice deficient for PKC δ ¹¹. In searching for the mechanism of the disease, we found that PKC δ -deficiency renders B cells BAFF-independent. Injection of a soluble BAFF receptor-Fc decoy, BAFF-R:Fc¹², in PKC δ ^{+/+} mice reduces the cellularity of the lymphoid organs (Fig. 1a, upper panel) by emptying peripheral

B-cell compartments and blocking most B-cell maturation at the transitional 1 (T1) ($\text{IgM}^+\text{CD21}^-$) stage (Fig. 1b, upper panel). Unlike $\text{PKC}\delta^{+/+}$ mice, injection of the soluble BAFF receptor into $\text{PKC}\delta$ -deficient mice has no effect on either the size of the B-cell compartment or B-cell survival and maturation (Fig. 1a and b, upper panels).

The independence of $\text{PKC}\delta^{-/-}$ B cells on BAFF was confirmed by analysis of B-cell subpopulations in mice in which $\text{PKC}\delta$ -deficiency is combined with a loss-of-function mutation in the BAFF receptor (BAFF-R). To produce these mice, the $\text{PKC}\delta^{-/-}$ mice were bred to A/WySnJ mutant mice that express a signalling-incompetent form of the BAFF-R^{13,14}. In the presence of $\text{PKC}\delta$, lack of BAFF signalling leads to reduction in the size of the lymphoid organs (Fig. 1a, lower panel), a decline in peripheral B-cell numbers (Fig. 1c) and a partial block in the maturation of immature ($\text{IgM}^+\text{CD21}^-$) to mature ($\text{IgM}^+\text{CD21}^+$) B cells (Fig. 1b, lower panel). In contrast to $\text{PKC}\delta^{+/+}$ A/WySnJ mice, $\text{PKC}\delta^{-/-}$ A/WySnJ mice have even more splenic and lymph node B cells and display a normal pattern of peripheral B-cell maturation (Fig. 1b and c, and data not shown). The recovery of the peripheral B-cell compartment in the absence of $\text{PKC}\delta$ cannot be attributed to higher expression of BAFF-R as $\text{PKC}\delta$ -deficient B cells have wild-type levels of BAFF-R messenger RNA (Supplementary Fig. 1a) and BAFF levels in the serum of the $\text{PKC}\delta^{-/-}$ mice are

within $\text{PKC}\delta^{+/+}$ limits (Supplementary Fig. 1b).

The greater than wild-type size of the $\text{PKC}\delta^{-/-}$ B-cell compartment in the absence of BAFF signalling could be due to the increased lifespan of $\text{PKC}\delta$ -deficient B cells *in vitro* and *in vivo* (Fig. 2a and c). Most $\text{PKC}\delta^{+/+}$ B cells cultured in serum-containing medium die within 3–4 days (Fig. 2a). $\text{PKC}\delta$ -deficiency renders B cells less susceptible to spontaneous death; about 20% remain alive even on day 14 of culture (Fig. 2a). Increased survival of the $\text{PKC}\delta$ -deficient B cells is not due to increased production of pro-survival cytokines such as interleukin-6 (IL-6). In the absence of stimulation, both $\text{PKC}\delta^{+/+}$ and $\text{PKC}\delta$ -deficient B cells express equal levels of IL-6 mRNA, and secrete similar amounts of IL-6 (Supplementary Fig. 2a). Furthermore, *in vitro* depletion of IL-6 with a neutralizing anti-IL-6 antibody does not affect the survival of $\text{PKC}\delta$ -deficient B cells (Supplementary Fig. 2b). The $\text{PKC}\delta$ -deficient B cells that survive *in vitro* for 14 days are functional and can be activated to proliferate by anti-IgM with IL-4, bacterial lipopolysaccharide (LPS), or unmethylated oligonucleotide (CpG) with kinetics similar to activation of freshly isolated $\text{PKC}\delta^{+/+}$ or $\text{PKC}\delta$ -deficient B cells (Fig. 2b).

It is possible that $\text{PKC}\delta$ controls B-cell survival in a BAFF-independent fashion and survival of $\text{PKC}\delta$ -deficient B cells reflects the general pro-apoptotic activity of $\text{PKC}\delta$. To test this model, we generated mice in which $\text{PKC}\delta$ -deficiency is combined with Btk-

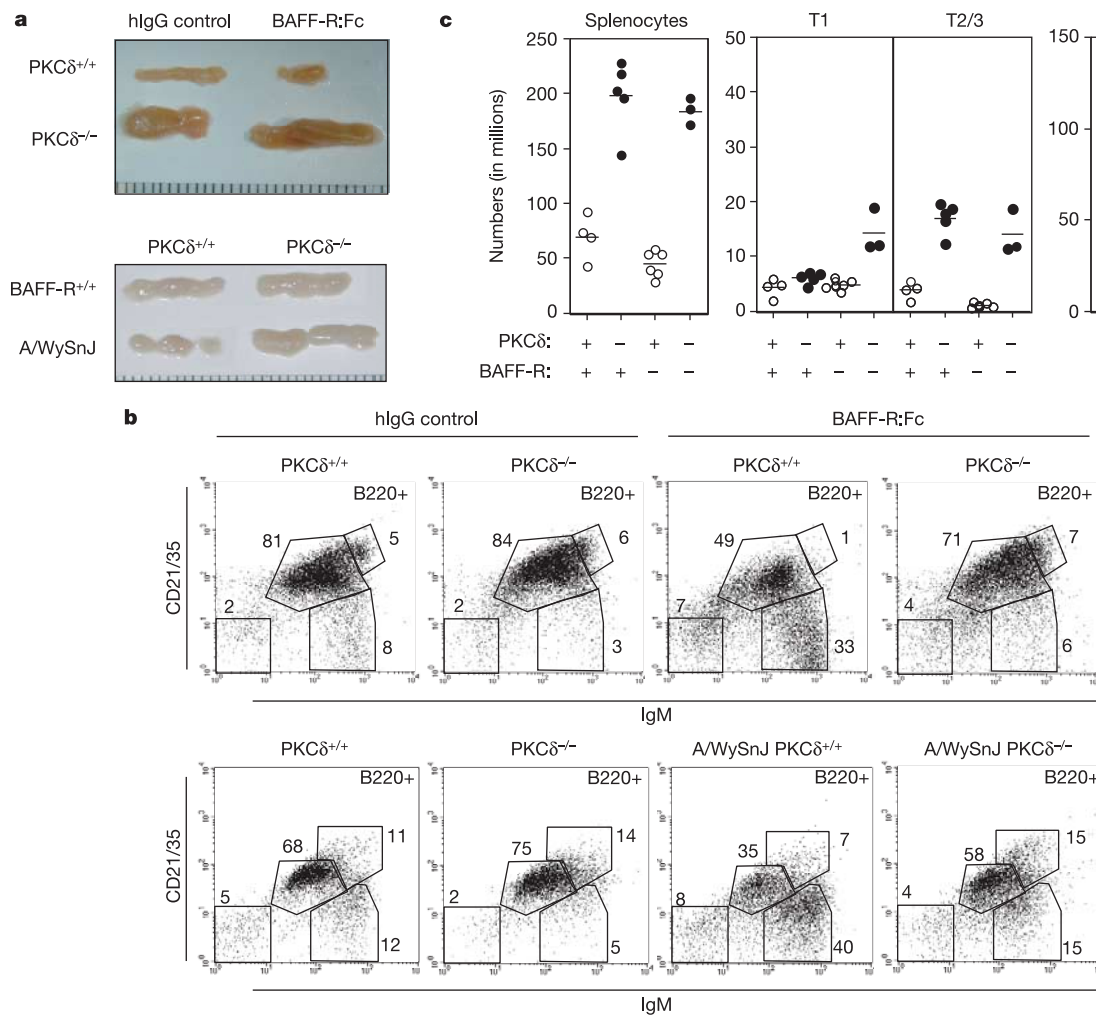


Figure 1 $\text{PKC}\delta$ -deficient B cells are independent of BAFF. **a**, Lymph nodes of mice injected with soluble BAFF-R:Fc (upper panel) or of 8-week-old A/WySnJ $\text{PKC}\delta^{-/-}$ or littermate control mice (lower panel) are shown. **b**, **c**, $\text{PKC}\delta$ -deficiency releases the block in B-cell maturation in the absence of BAFF signalling. **b**, Maturity of splenic B cells was assessed by analysis of the surface expressed CD21/35 and IgM. Data are representative

of three independent experiments. **c**, Absolute numbers of splenic CD19^+ immature ($\text{IgM}^+\text{CD21}^-$) and mature ($\text{IgM}^+\text{CD21}^+$) B cells of A/WySnJ $\text{PKC}\delta^{+/+}$ (open symbols) or A/WySnJ $\text{PKC}\delta^{-/-}$ (filled symbols) mice are shown. Each dot represents an individual mouse.

deficiency¹⁵. The latter mutation causes a major decline in the survival of naïve B cells *in vitro*; this can be reversed by over-expression of the Bcl-2 protein¹⁶. Our data show that PKC δ -deficiency fails to rescue poor survival (Fig. 2d) and B-cell cellularity in the lymphoid organs of the PKC δ ^{-/-} Btk^{-/-} mice (Fig. 2e and data not shown). The existence of a PKC δ -dependent pro-apoptotic pathway that operates via potentially novel signalling mechanisms is also supported by Affymetrix gene chip and RT-PCR analyses, which show no significant differences in the PKC δ -deficient and PKC δ ^{+/+} B cells in expression of mRNAs encoding known anti- or pro-apoptotic proteins (Supplementary Fig. 3a). Furthermore, deficiency in PKC δ does not affect BAFF-mediated processing of the NF- κ B2 that was previously implicated in positive regulation of B-cell survival^{17,18} (Supplementary Fig. 3b). Collectively, these results exclude a general anti-apoptotic effect of PKC δ -deficiency and suggest a specific BAFF-dependent pathway that regulates B-cell survival by negative regulation of PKC δ .

Previous studies on the role of PKC δ in cell death revealed a pro-apoptotic function of this enzyme in cells of various types¹⁹. It was shown that initiation of apoptosis caused by various agents correlates with nuclear translocation of the full-length PKC δ

(PKC δ -FL) or its catalytic fragment (PKC δ -CF), generated by caspase-3-mediated cleavage at the site that separates the regulatory and the catalytic subunits of the enzyme^{20,21}. The pro-apoptotic activity of PKC δ requires nuclear localization²² and correlates directly with its enzymatic activity; this is demonstrated by differences in the efficiency of cell death induction by expression of exogenous PKC δ -FL and PKC δ -CF, which possess higher enzymatic activity than non-cleaved PKC δ -FL²³. These findings, as well as the failure of catalytically inactive PKC δ to cause cell death²², suggest that the pro-apoptotic function of PKC δ is mediated through phosphorylation of nuclear targets.

In *ex vivo* isolated splenic B cells, the majority of the PKC δ protein is located in the cytosol (Fig. 3a). However, 24 h incubation of B cells in serum-containing medium, which triggers some of the B cells to death (Fig. 2a), leads to an increase in nuclear PKC δ -FL (Fig. 3a). This process is independent of the activity of caspases that are implicated in cell death. Incubation of *ex vivo* isolated B cells with the pan-caspase inhibitor Z-VAD-FMK or with inhibitors specific for caspase-3, -6 or -9 does not prevent nuclear accumulation of PKC δ -FL (Fig. 3b and data not shown). Whereas the role of caspases in cell death is obvious, our data suggest that nuclear

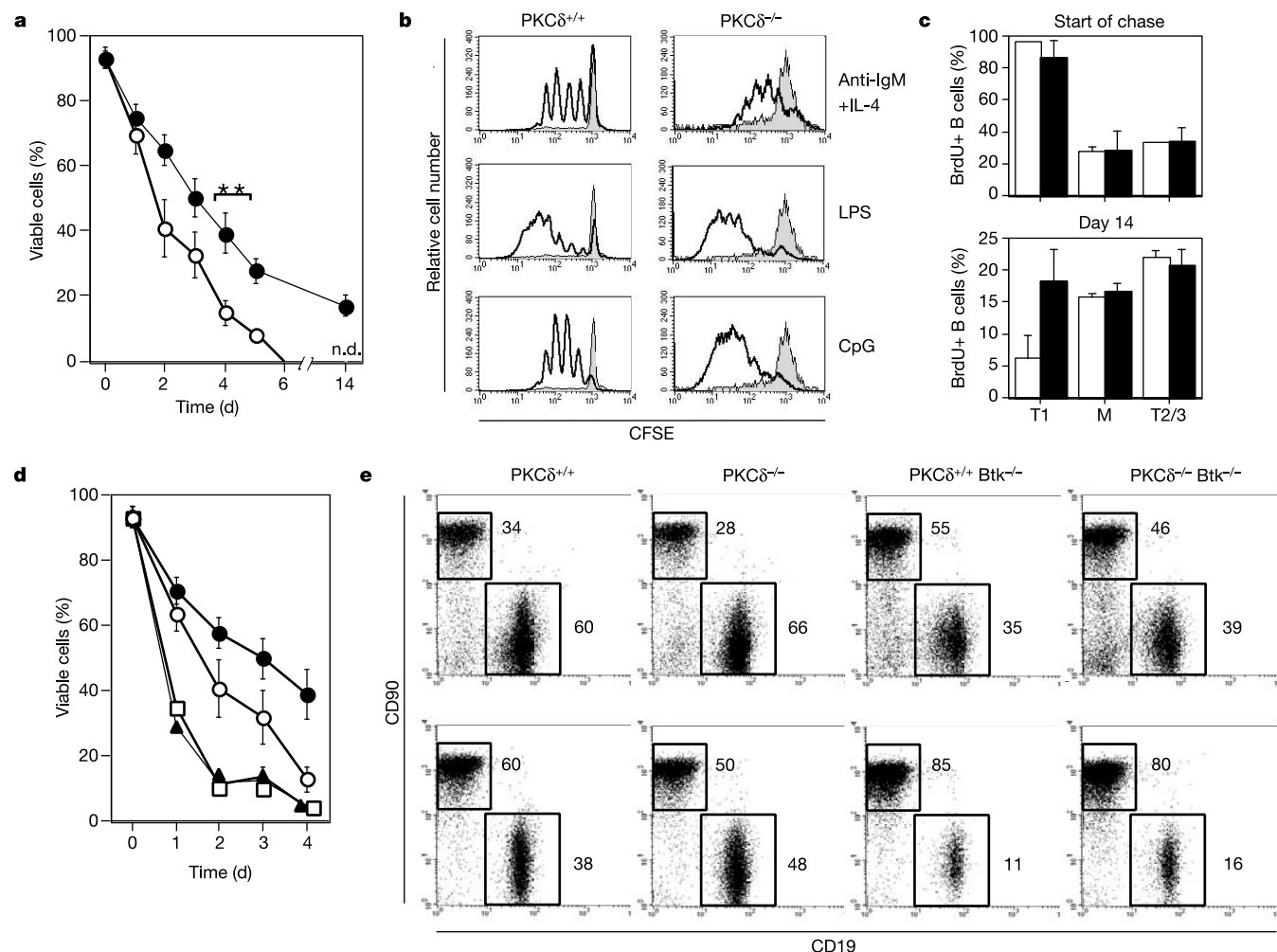


Figure 2 PKC δ promotes B-cell survival. **a**, B-cell survival *in vitro*. Frequencies of viable PKC δ ^{+/+} (open circles) or PKC δ ^{-/-} (filled circles) B cells incubated in medium. Data are means $\pm$ s.d. from 5–10 independent experiments. n.d., not detectable. **P < 0.0001 by t test. **b**, Proliferation of freshly isolated PKC δ ^{+/+} or PKC δ ^{-/-} B cells (day 14 after culture) in the absence (grey histogram) or presence of stimuli (thick line) was measured by CFDA-SE labelling. **c**, B-cell lifespan *in vivo* was measured by BrdU incorporation. Open bars, PKC δ ^{+/+}; filled bars, PKC δ ^{-/-}. Data are means $\pm$ d.f. from three mice.

d and **e**, PKC δ -deficiency does not rescue the survival of Btk-deficient B cells. **d**, Frequencies of viable B cells after incubation in medium. Open circles, PKC δ ^{+/+} Btk^{+/+}; filled circles, PKC δ ^{-/-} Btk^{+/+}; open squares, Btk^{-/-}; filled triangles, PKC δ ^{-/-} Btk^{-/-}. Data are means $\pm$ s.d. from two independent experiments. **e**, Frequencies of splenic (upper panel) and lymph node (lower panel) T (CD90⁺) and B (CD19⁺) cells.

translocation of PKC δ and hence related to it B-cell death, are caspase-independent. It is possible that BAFF regulates B-cell survival through negative regulation of nuclear PKC δ . Indeed, we found that treatment of B cells with recombinant BAFF prevents nuclear accumulation of PKC δ (Fig. 3a).

The connection between PKC δ translocation to the nucleus and cell death raises the question of nuclear targets for PKC δ . Within the nucleus, the most abundant PKC substrates are histones. It has been shown that phosphorylation of histone H2B at serine 14 (S14-H2B)

is associated with cell death^{24,25}. We therefore hypothesized that the ability of PKC δ to induce apoptosis might reflect its ability to phosphorylate S14 of the histone H2B.

Expression of the exogenous PKC δ catalytic fragment (PKC δ -CF) in NIH3T3 cells and in PKC $\delta^{+/+}$ or PKC δ -deficient embryonic fibroblasts increases S14-H2B phosphorylation to levels higher than those observed in non-transfected cells (or in cells transfected with either the full-length PKC δ or a PKC δ variant that lacks the caspase-cleavage site) (Fig. 4a and b). The overall higher levels of S14-H2B phosphorylation in NIH3T3 cells that express various PKC δ variants is likely to reflect the higher expression levels of the PKC δ -GFP fusion proteins in NIH3T3 cells as compared to embryonic fibroblasts. In lymphoid cells S14-H2B appears to be a specific substrate for PKC δ , as neither phosphorylation of serine 10 (S10-H3) nor serine 28 (S28-H3) of histone H3 is affected by expression of highly pro-apoptotic PKC δ -CF (Fig. 4c). Involvement of PKC δ in S14-H2B phosphorylation is further supported by the analysis of *ex vivo* isolated peripheral resting B cells (Fig. 4d). The S14-H2B phosphorylation could be detected in freshly isolated PKC $\delta^{+/+}$ B cells that do not contain nuclear PKC δ . Hence kinases other than PKC δ can also contribute to S14-H2B phosphorylation in B cells. One of them could be the *sterile twenty kinase* Mst1 that can phosphorylate S14-H2B in mammalian cells²⁵. Nonetheless, the significantly reduced levels of S14-H2B phosphorylation in PKC δ -deficient B cells points to a major contribution of PKC δ to regulation of S14-H2B phosphorylation in resting B cells (Fig. 4d). Appearance of phosphorylated S14-H2B is unlikely to be the consequence of apoptosis as such. Thus apoptotic death of PKC $\delta^{+/+}$ B cells induced by anti-Fas antibody is not associated with an increase in S14-H2B phosphorylation (Fig. 4e). This result, along with our data on unaltered PKC $\delta^{-/-}$ B-cell sensitivity to Fas-mediated death¹¹, point to a specific role of PKC δ and S14-H2B phosphorylation in spontaneous B-cell death.

Whereas our data reveal an important role of PKC δ in regulation of S14-H2B phosphorylation, the role of S14-H2B phosphorylation in B-cell death remains unclear. The presence of multiple H2B genes

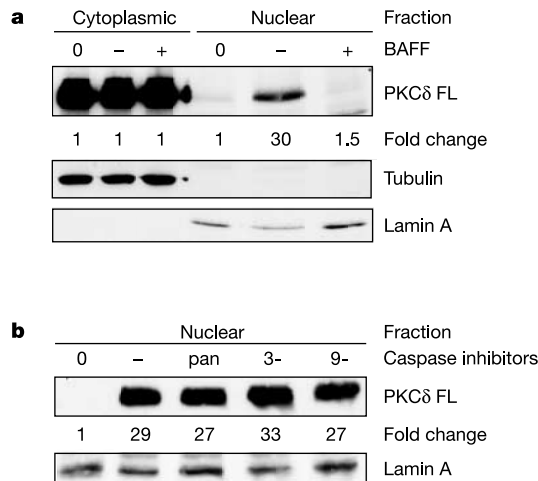


Figure 3 Nuclear expression of PKC δ in *ex vivo* isolated B cells is negatively regulated by BAFF. Purified B cells were incubated for 24 h in serum-containing medium in the absence or presence of BAFF (**a**), or with pan-, caspase-3 or -9-specific inhibitors (**b**), followed by preparation of cytosolic and nuclear extracts. Expression of PKC δ was measured by western blot analysis of extracts using anti-PKC δ antibody. Protein loading was controlled by tubulin and lamin A expression analysis. 0, freshly isolated cells. Data were quantified by NIH Image software.

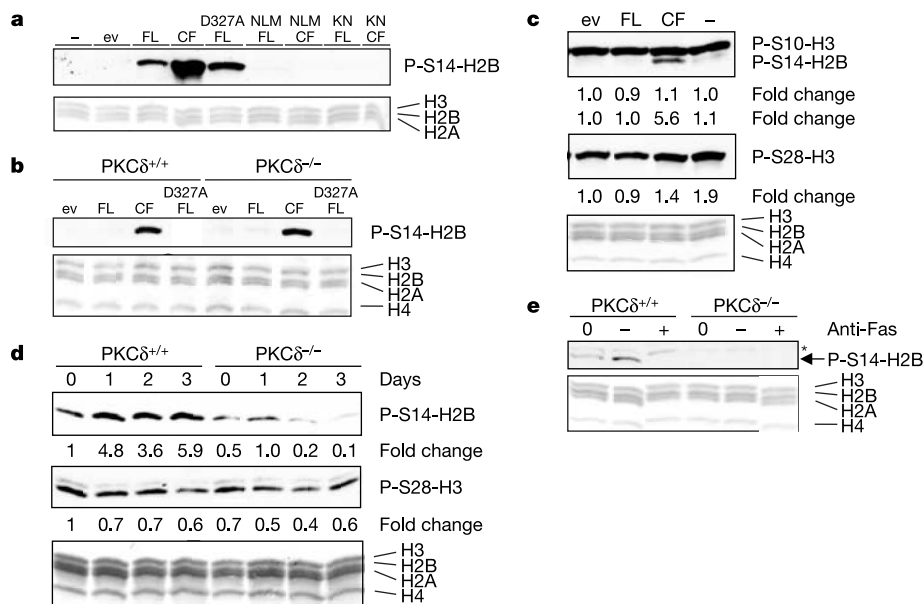


Figure 4 PKC δ contributes to cell-death-related phosphorylation of S14-H2B. **a–c**, The levels of histone phosphorylation were measured by western blot analysis using phospho- and residue-specific antibodies in NIH3T3 cells (**a**); mouse embryonic fibroblasts generated from PKC $\delta^{+/+}$ and PKC $\delta^{-/-}$ mice (**b**); murine A20 cells (**c**) overexpressing full-length (FL), catalytic fragment (CF), caspase-cleavage-deficient (D327A), nuclear localization sequence (NLM)-deficient or kinase-inactive (KN) PKC δ . Protein loading was controlled by Ponceau staining of core histones. -, not infected; ev, empty vector.

d, Purified PKC $\delta^{+/+}$ and PKC $\delta^{-/-}$ B cells were incubated in serum-containing medium for different time periods. The levels of S14-H2B and S28-H3 phosphorylation were measured by western blot analysis. Data were quantified by NIH Image software.

e, Not-activated (-) or anti-CD40-triggered (+) purified PKC $\delta^{+/+}$ and PKC $\delta^{-/-}$ splenic B cells were subjected to anti-Fas antibody treatment for 6 h. The level of S14-H2B phosphorylation was measured by western blot analysis. 0, freshly isolated cells.

in the mouse genome precludes analysis of the physiological role of S14-H2B phosphorylation by mutagenesis *in vivo*. Nonetheless, attenuation of hydrogen-peroxide-induced cell death by replacement of S14-H2B of *Saccharomyces cerevisiae* (S. H. Ahn and D. Allis, personal communication), and recent data that show accumulation of the S14-H2B at sites of DNA breaks²⁶, indicate the importance of P-S14-H2B in apoptosis. Should S14-H2B phosphorylation determine the rate of spontaneous B-cell death, then the lower levels of S14-H2B phosphorylation in the absence of PKC δ may explain slower (compared to PKC $\delta^{+/+}$ B cells) kinetics of PKC δ -deficient B cell death *in vitro*.

In conclusion, our results show that BAFF promotes B-cell survival not only through the NF- κ B-mediated increase in the expression of anti-apoptotic proteins, but also by actively preventing nuclear translocation of PKC δ . □

Methods

Mice

Mice with PKC δ -deficiency¹¹, initially on 129/Sv*Ola background, were backcrossed to C57BL/6 mice for at least 10 generations. PKC $\delta^{+/+}$ mice were intercrossed to generate mutant and control mice. The predominant C57BL/6 genetic background of PKC $\delta^{-/-}$ mice was verified by microsatellite genotyping performed at the Rockefeller University Genotyping Resource Center (Supplementary Table 1). BAFF-R mutant PKC $\delta^{-/-}$ or Btk $^{-/-}$ PKC δ -doubly-deficient mice and the corresponding littermate control mice were generated by crossing either PKC δ -deficient and A/WySnJ mice (The Jackson Laboratory) or PKC δ -deficient and Btk $^{-/-}$ mice (The Jackson Laboratory) and intercrossing the resulting F1 mice to generate F2. For neutralization of BAFF, 30-week-old littermate PKC δ -deficient or PKC $\delta^{+/+}$ were injected i.p. with 100 μ g of control-Ig (human IgG, Novartis) or with BAFF-R:Fc¹² three times a week for five weeks. For 5-bromo-2-deoxyuridine (BrdU) labelling, PKC $\delta^{+/+}$ and PKC δ -deficient mice received 1 mg ml⁻¹ of BrdU in their drinking water for 14 days. BrdU incorporation into B cells was determined by staining with FITC-labelled anti-BrdU antibody (BD Pharmingen). Frequency of BrdU⁺ B cells was measured by flow cytometry.

Cell lines, transient transfections, nuclear extract preparation

Mouse embryonic fibroblasts were generated from 13.5-day PKC $\delta^{+/+}$ and PKC $\delta^{-/-}$ embryos by standard procedure. Transfection of NIH3T3, Jurkat, BJAB, Ramos and A20 cells was done either according to the Lipofectamine Plus kit (Invitrogen) or by electroporation. Seventy-two hours after transfection with various constructs²², cells were washed with PBS and lysed as described²⁷. Briefly, cells were lysed for 15 min on ice in 320 mM sucrose, 10 mM Tris pH 8.0, 3 mM CaCl₂, 2 mM magnesium acetate, 0.1 mM EDTA, 0.5% NP-40, supplemented with protease (Sigma) and phosphatase (Calbiochem) inhibitor cocktails. After 5 min of centrifugation at 500g, supernatants were saved as the cytosolic fraction. The nuclei were washed in the same buffer without NP-40 and then lysed in 1% NP-40, 500 mM NaCl, 50 mM Tris pH 8.0, 10% glycerol, and sonicated. The supernatants after 15 min of centrifugation at 16×10^3 g were used as nuclear fraction.

Immunoblotting analysis

Western blot analysis was performed by separating equal amounts of protein extracts by SDS-PAGE and blotting to nitrocellulose membrane by standard procedures. For immunoblotting of cytoplasmic and nuclear PKC δ , protein extracts of purified B cells were analysed by immunoblotting with an antibody against PKC δ (Santa Cruz). Protein loading was controlled with anti-lamin A antibody (Cell Signaling) and anti-tubulin antibody (Sigma). The expression of p100 and p52 of NF- κ B2 was studied using protein extracts prepared as described¹⁸ and antibodies against NF- κ B2 (Upstate) or β -actin (Oncogene). For immunoblotting of nuclear histones, protein loading was controlled by reversible staining of the core histones H3, H2B, H2A and H4 on the membrane with Ponceau S solution (Sigma), followed by probing with phospho-specific antibodies against S14-H2B, S10-H3 and S28-H3 (Upstate). Data were quantified by NIH Image software.

B-cell purification, proliferation, staining and flow cytometry

Splenic B cells were purified as described previously¹¹ and stimulated *in vitro* with 2.4 μ g ml⁻¹ F(ab')₂ fragment of goat anti-mouse IgM (Jackson ImmunoResearch) in combination with 25 U ml⁻¹ recombinant mouse IL-4 (R&D), 5 μ g ml⁻¹ bacterial LPS (Sigma), or 10 nM CpG ODN 1638 (MWG Biotech). For neutralization of any IL-6 bioactivity, purified rat anti-mouse IL-6 (MP5-20F3) and rat IgG1 isotype-matched control (R3-34) antibodies (BD Pharmingen) were used. The caspase inhibitors Z-VAD-FMK, Z-VEID-FMK, Z-LEHD-FMK (BioVision) and DEVD-CHO (Biosource) were used at a concentration of 100 nM. For induced apoptosis, purified B cells were activated with 0.6 μ g ml⁻¹ anti-CD40 antibody for 12 h and then subjected to 50 ng ml⁻¹ anti-Fas antibody treatment. Labelling of cells with 5-(and 6-)-carboxyfluorescein diacetate, succinimidyl ester (CFDA-SE; Molecular Probes) for analysis of proliferation was performed as described¹¹. The decline in CFSE fluorescence as a measure of B-cell proliferation was determined by FACS analysis. Lymphocyte surface marker expression was analysed using monoclonal antibodies against CD19, CD21/35, CD23, CD90, B220 (BD Pharmingen) or F(ab')₂ fragment of goat anti-mouse IgM (Jackson ImmunoResearch) as described previously¹¹.

Construction and generation of MigRI-PKC δ retroviruses

Construction of the GFP-tagged PKC δ -FL, its catalytic fragment (PKC δ -CF), caspase-cleavage mutant (D327A), or PKC δ with mutated nuclear localization sequence (NLM), or dominant-negative (K376R) PKC δ were described previously²². The PKC δ -GFP fusions were excised with NotI and XhoI, blunt-ended and ligated into the MigRI vector²⁸ after removing the IRES-GFP. The infectious retroviral particles were then produced by cotransfecting retrovirus vector DNA with the helper plasmid pCL-Eco into Bosc23 packaging cells. Retroviral supernatants were collected and filtered 48 h later. The PKC $\delta^{+/+}$ and PKC $\delta^{-/-}$ MEFs were then infected with the supernatant with 8 μ g ml⁻¹ polybrene (Sigma). Aliquots of cells were either subjected to FACS analysis for GFP expression or lysed 48 h post-infection.

Cell survival assay

Purified B cells were cultured either in medium alone or in the presence of 500 ng ml⁻¹ of recombinant BAFF²⁹ for the indicated time and stained with Annexin V (Roche) and 7-aminoactinomycin D (7-AAD; Sigma-Aldrich)³⁰.

Received 12 July; accepted 16 August 2004; doi:10.1038/nature02955.

Published online 8 September 2004.

- Wardemann, H. et al. Predominant autoantibody production by early human B cell precursors. *Science* **301**, 1374–1377 (2003).
- Goodnow, C. C. et al. Altered immunoglobulin expression and functional silencing of self-reactive B lymphocytes in transgenic mice. *Nature* **334**, 676–682 (1988).
- Nemazee, D. A. & Burki, K. Clonal deletion of B lymphocytes in a transgenic mouse bearing anti-MHC class I antibody genes. *Nature* **337**, 562–566 (1989).
- Erikson, J. et al. Expression of anti-DNA immunoglobulin transgenes in non-autoimmune mice. *Nature* **349**, 331–334 (1991).
- Hartley, S. B. et al. Elimination from peripheral lymphoid tissues of self-reactive B lymphocytes recognizing membrane-bound antigens. *Nature* **353**, 765–769 (1991).
- Goodnow, C. C. Balancing immunity and tolerance: deleting and tuning lymphocyte repertoires. *Proc. Natl Acad. Sci. USA* **93**, 2264–2271 (1996).
- Mackay, F. & Kalled, S. L. TNF ligands and receptors in autoimmunity: an update. *Curr. Opin. Immunol.* **14**, 783–790 (2002).
- Mackay, F. et al. Mice transgenic for BAFF develop lymphocytic disorders along with autoimmune manifestations. *J. Exp. Med.* **190**, 1697–1710 (1999).
- Gross, J. A. et al. TACI and BCMA are receptors for a TNF homologue implicated in B-cell autoimmune disease. *Nature* **404**, 995–999 (2000).
- Khare, S. D. et al. Severe B cell hyperplasia and autoimmune disease in TALL-1 transgenic mice. *Proc. Natl Acad. Sci. USA* **97**, 3370–3375 (2000).
- Mecklenbrauker, I., Saijo, K., Zheng, N. Y., Leitges, M. & Tarakhovsky, A. Protein kinase C δ controls self-antigen-induced B-cell tolerance. *Nature* **416**, 860–865 (2002).
- Pelletier, M. et al. Comparison of soluble decoy IgG fusion proteins of BAFF-R and BCMA as antagonists for BAFF. *J. Biol. Chem.* **278**, 33127–33133 (2003).
- Thompson, J. S. et al. BAFF-R, a newly identified TNF receptor that specifically interacts with BAFF. *Science* **293**, 2108–2111 (2001).
- Yan, M. et al. Identification of a novel receptor for B lymphocyte stimulator that is mutated in a mouse strain with severe B cell deficiency. *Curr. Biol.* **11**, 1547–1552 (2001).
- Khan, W. N. et al. Defective B cell development and function in Btk-deficient mice. *Immunity* **3**, 283–299 (1995).
- Woodland, R. T., Schmidt, M. R., Korsmeyer, S. J. & Gravel, K. A. Regulation of B cell survival in xid mice by the proto-oncogene bcl-2. *J. Immunol.* **156**, 2143–2154 (1996).
- Claudio, E., Brown, K., Park, S., Wang, H. & Siebenlist, U. BAFF-induced NEMO-independent processing of NF- κ B2 in maturing B cells. *Nature Immunol.* **3**, 958–965 (2002).
- Kayagaki, N. et al. BAFF/BLyS receptor 3 binds the B cell survival factor BAFF ligand through a discrete surface loop and promotes processing of NF- κ B2. *Immunity* **17**, 515–524 (2002).
- Mischak, H. et al. Overexpression of protein kinase C- δ and - ϵ in NIH3T3 cells induces opposite effects on growth, morphology, anchorage dependence, and tumorigenicity. *J. Biol. Chem.* **268**, 6090–6096 (1993).
- Emoto, Y. et al. Proteolytic activation of protein kinase C δ by an ICE-like protease in apoptotic cells. *EMBO J.* **14**, 6148–6156 (1995).
- Ghayur, T. et al. Proteolytic activation of protein kinase C δ by an ICE/CED 3-like protease induces characteristics of apoptosis. *J. Exp. Med.* **184**, 2399–2404 (1996).
- DeVries, T. A., Neville, M. C. & Reyland, M. E. Nuclear import of PKC δ is required for apoptosis: identification of a novel nuclear import sequence. *EMBO J.* **21**, 6050–6060 (2002).
- Reyland, M. E., Anderson, S. M., Matassa, A. A., Barzen, K. A. & Quissell, D. O. Protein kinase C δ is essential for etoposide-induced apoptosis in salivary gland acinar cells. *J. Biol. Chem.* **274**, 19115–19123 (1999).
- Ajito, K. Histone H2B phosphorylation in mammalian apoptotic cells. An association with DNA fragmentation. *J. Biol. Chem.* **275**, 439–443 (2000).
- Cheung, W. L. et al. Apoptotic phosphorylation of histone H2B is mediated by mammalian sterile twenty kinase. *Cell* **113**, 507–517 (2003).
- Fernandez-Capetillo, O., Allis, C. D. & Nussenzweig, A. Phosphorylation of histone H2B at DNA double-strand breaks. *J. Exp. Med.* **199**, 1671–1677 (2004).
- Su, I. H. et al. Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement. *Nature Immunol.* **4**, 124–131 (2003).
- Van Parijs, L. et al. Uncoupling IL-2 signals that regulate T cell proliferation, survival, and Fas-mediated activation-induced cell death. *Immunity* **11**, 281–288 (1999).
- Karpusas, M. et al. Crystal structure of extracellular human BAFF, a TNF family member that stimulates B lymphocytes. *J. Mol. Biol.* **315**, 1145–1154 (2002).
- Saijo, K. et al. Protein kinase C β controls nuclear factor κ B activation in B cells through selective regulation of the I κ B kinase α . *J. Exp. Med.* **195**, 1647–1652 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. H. Ahn, D. Allis, M. Reyland, U. Siebenlist, the Rockefeller University Genotyping Resource Center and the MSKCC Genomics Core Laboratory for providing cells, vectors, reagents and technical assistance. We also thank E. Besmer for help with manuscript preparation, and A. Patke and members of the Tarakhovsky laboratory for discussions. This work was supported by The Irene Diamond Fund/Professorship Program (A.T.), the NIH (A.T.) and The S.L.E. Foundation (I.M.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.T. (tarakho@mail.rockefeller.edu).

NF- κ B functions as a tumour promoter in inflammation-associated cancer

Eli Pikarsky^{1*}, Rinnat M. Porat^{1*}, Ilan Stein^{1,2*}, Rinat Abramovitch³, Sharon Amit², Shafika Kasem¹, Elena Gutkovich-Pyest², Simcha Urieli-Shoval⁴, Eithan Galun³ & Yinon Ben-Neriah²

¹Department of Pathology, Hadassah-Hebrew University Medical Center, Jerusalem 91120, Israel

²The Lautenberg Center for Immunology, Hebrew University-Hadassah Medical School, Jerusalem 91120, Israel

³Goldyne Savad Institute of Gene Therapy, Jerusalem 91120, Israel

⁴Hematology Unit, Hadassah University Hospital, Mount Scopus, Jerusalem 91120, Israel

* These authors contributed equally to this work

The causes of sporadic human cancer are seldom recognized, but it is estimated that carcinogen exposure and chronic inflammation are two important underlying conditions for tumour development, the latter accounting for approximately 20% of human cancer¹. Whereas the causal relationship between carcinogen exposure and cancer has been intensely investigated², the molecular and cellular mechanisms linking chronic inflammation to tumorigenesis remain largely unresolved¹. We proposed that activation of the nuclear factor κ B (NF- κ B), a hallmark of inflammatory responses³ that is frequently detected in tumours^{4,5}, may constitute a missing link between inflammation and cancer. To test this hypothesis, we studied the *Mdr2*-knockout mouse strain, which spontaneously develops cholestatic hepatitis followed by hepatocellular carcinoma⁶, a prototype of inflammation-associated cancer⁷. We monitored hepatitis and cancer progression in *Mdr2*-knockout mice, and here we show that the inflammatory process triggers hepatocyte NF- κ B through upregulation of tumour-necrosis factor- α (TNF α) in adjacent endothelial and inflammatory cells. Switching off NF- κ B in mice from birth to seven months of age, using a hepatocyte-specific inducible I κ B-super-repressor transgene, had no effect on the course of hepatitis, nor did it affect early phases of hepatocyte transformation. By contrast, suppressing NF- κ B inhibition through anti-TNF α treatment or induction of I κ B-super-repressor in later stages of tumour development resulted in apoptosis of transformed hepatocytes and failure to progress to hepatocellular carcinoma. Our studies thus indicate that NF- κ B is essential for promoting inflammation-associated cancer, and is therefore a potential target for cancer prevention in chronic inflammatory diseases.

Hepatocellular carcinoma (HCC), the third leading cause of cancer mortality worldwide, commonly develops in the background of chronic hepatitis⁸. We confirmed and extended previous

results^{6,9}, according to which tumour development in the *Mdr2*-knockout mouse, similarly to human HCC⁸, progresses through distinct phases: inflammation, dysplasia, dysplastic nodules (adenoma-like), carcinoma and metastasis (Supplementary Fig. 1a, b). Hepatitis is the earliest sign of disease in the *Mdr2*-knockout mice, and is characterized by an extensive periductular and periportal mixed inflammatory infiltrate, rich in CD3-positive cells (Supplementary Fig. 1c).

NF- κ B activation is often observed in human HCC, particularly following hepatitis⁸. The possibility that NF- κ B activation is involved in *Mdr2*-knockout hepatocarcinogenesis was investigated by RelA (p65) immunostaining (Fig. 1a). Nuclear staining, indicative of NF- κ B activation, was evident in knockout livers of mice at all ages, but not in normal, age-matched mice; this prompted a search for the source of NF- κ B activation in the knockout mice. In some neoplasms such as Hodgkin's disease and certain lymphomas, NF- κ B activation is an intrinsic, tumour autonomous feature, possibly related to mutations in its signalling pathway¹⁰. If a similar mechanism occurs in *Mdr2*-knockout HCC, one would observe nuclear NF- κ B activation in all neoplastic hepatocytes. However, this was not the case; NF- κ B activation appeared scattered at the parenchyma, predominantly adjacent to inflamed portal tracts (Fig. 1e, left panel), suggesting an inflammation-induced phenomenon. To evaluate this possibility, we fed knockout and wild-type mice with the non-steroidal anti-inflammatory drug (NSAID) ibuprofen for 10 days. This treatment resulted in decreased inflammation, evident by histological analysis (Fig. 1b), decreased serum alanine aminotransferase (ALT), a marker for hepatocyte damage, (Fig. 1c) and fewer neutrophils (Fig. 1c, d). p65 immunostaining showed that the degree of hepatocyte NF- κ B activation in the NSAID-treated group was significantly lower than in the controls (Fig. 1e). In contrast, ibuprofen had no effect on hepatocyte NF- κ B activation following intraperitoneal TNF α administration (data not shown). Hence, it seems that NF- κ B activation in the knockout hepatocytes is secondary to parenchymal infiltration by inflammatory cells.

One likely mediator of NF- κ B activation in the inflamed liver is TNF α , particularly its membrane form, which is upregulated in knockout livers (Fig. 2a). To study this possibility, we treated knockout mice with neutralizing anti-TNF α antibodies for three days and analysed hepatocyte NF- κ B activation. Anti-TNF α treatment abolished hepatocyte p65 nuclear staining (Fig. 2b), confirming the prediction that NF- κ B activation is primarily induced by TNF α . To determine the source of TNF α in the inflamed liver, we fractionated wild-type and knockout livers to hepatocytes (>80% purity by haematoxylin-and-eosin (H&E)) and other cells (<2% hepatocytes). Analysis of the two fractions for TNF α expression by real-time polymerase chain reaction (PCR) confirmed that the source of TNF α was the non-hepatocyte fraction, which expressed fivefold more TNF α than the corresponding fraction of wild-type mice (Fig. 2c).

To study the relationship between the TNF α -producing cells and NF- κ B activation in the hepatocytes, liver sections were stained for both TNF α and p65 (Fig. 2d, f). TNF α expression was detected primarily at portal spaces, both in endothelial and inflammatory cells (Fig. 2d). Hepatocyte nuclear p65 staining was particularly abundant close to TNF α -positive portal spaces (Fig. 2f), indicating that hepatocyte NF- κ B is activated through paracrine TNF α stimulation. To confirm the identity of the TNF α producers and distinguish them from TNF α -binding cells, liver sections of wild-type, knockout and ibuprofen-treated mice were analysed by *in situ* messenger RNA hybridization. Similar to the immunostaining data, TNF α mRNA was found to be highly expressed by endothelial and inflammatory cells of knockout mice, but far less in wild-type and ibuprofen-treated knockout mice. TNF α expression was hardly detected in either knockout or wild-type hepatocytes (Fig. 2e).

The frequent activation of NF- κ B in knockout hepatocytes

suggested that inflammation-associated NF- κ B activation promotes neoplastic growth^{4,5}. To directly assess the role of NF- κ B in hepatocarcinogenesis, we bred *Mdr2*-knockout mice with ΔN -*I κ B*^{hep} mice, which carry two transgenes: a non-degradable ΔN -*I κ B* controlled by a tetracycline (tet)-regulated promoter and the tetracycline trans-activator under control of the hepatocyte specific C/EBP β promoter (TA^{LAP}1)¹¹. The tet-regulated promoter also directs the expression of luciferase; consequently, transgene expression levels and its tissue specificity can be determined using a live-mouse recording of luciferase activity¹¹. Crossing the bi-transgenic ΔN -*I κ B*^{hep} mice with *Mdr2*-knockout mice generates *Mdr2*^{-/-} ΔN -*I κ B*^{hep} (double mutant) mice amenable to NF- κ B modulation. Immunohistochemical staining of livers from the same animals with anti-luciferase antibodies demonstrated that only hepatocytes, but not the bile duct epithelium, express the transgene (Supplementary Fig. 2a). As expected, p65 immunostaining detected many positive hepatocyte nuclei in knockout and doxycycline (Dox)-treated (transgene-suppressed) mice, but none in the untreated double mutant mice (Supplementary Fig. 2b, c). Positively-stained biliary and Kupffer cells were detectable in all knockout and double mutant mice irrespective of Dox treatment (not shown), attesting to the specificity and efficiency of the transgene in blocking NF- κ B activation.

Because the origin of liver inflammation in the *Mdr2*-knockout is the biliary system⁶, it should not be affected by the hepatocyte-

specific ΔN -*I κ B*^{hep} transgene. Nevertheless, before assessing the role of NF- κ B in tumorigenesis, we wanted to rule out any possible effect of NF- κ B inhibition on the inflammatory process. Both the knockout and the double mutant mice had higher ALT levels than those of age-matched wild-type mice (Fig. 3b). Histological analysis of livers from both groups at all ages revealed comparable non-suppurative cholangitis with portal expansion due to ductular proliferation, a mixed portal inflammatory infiltrate and mild to moderate fibrosis (Fig. 3a). Both strains had high numbers of CD3-positive and myeloperoxidase (MPO)-positive cells in the portal tracts and the liver parenchyma. Thus, it seems that the fundamental inflammatory process in *Mdr2*-knockout mice is maintained in the double mutant mice and is independent of hepatocyte NF- κ B activity.

Mdr2-knockout hepatocytes are distinguishable from wild-type cells by several abnormal features: high proliferation rate, accelerated hyperplasia and dysplasia. Whereas the first are also characteristics of liver damage and partial hepatectomy¹², dysplasia is a premalignant condition¹³. Hepatocyte proliferation was evaluated by bromo-deoxy uridine (BrdU) incorporation and Ki-67 antigen staining. Both markers were clearly enhanced in knockout and double mutant mice compared with wild-type mice (Fig. 3b). In agreement with previous studies^{14,15}, NF- κ B deficiency doesn't compromise hepatocyte proliferation: there was no significant difference in either hepatocyte BrdU incorporation or Ki-67 staining between the knockout and the double mutant mice at any age.

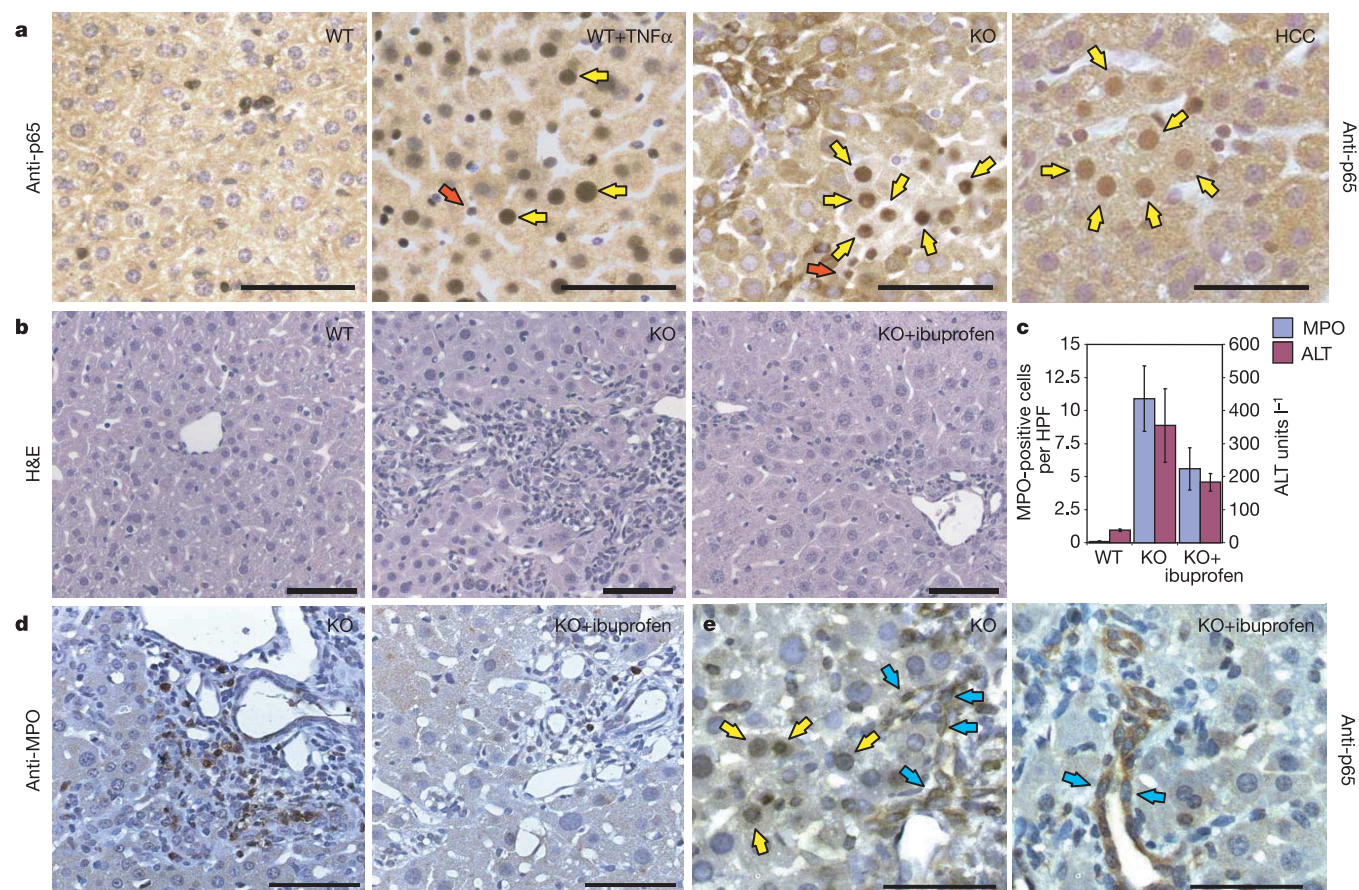


Figure 1 NF- κ B activation, a prominent component of *Mdr2*-knockout hepatitis, is abrogated by NSAID treatment. **a**, Tissue sections from wild-type and knockout mice were immunostained with antibodies against p65/RelA. TNF α -treated wild-type liver serves as a positive control for p65 staining. HCC: a section from a hepatocellular carcinoma that developed in a knockout mouse. **b**, H&E-stained sections from livers of 2.5-month-old wild-type, knockout and ibuprofen-treated knockout mice. Note decreased inflammatory infiltrate after ibuprofen treatment. **c**, MPO cell counts and serum ALT levels from

wild-type, knockout and ibuprofen-treated knockout mice (mean $\pm$ s.e.m.). **d-e**, Immunohistochemical stains for MPO (**d**) and p65 (**e**) were performed on sections from untreated and ibuprofen-treated knockout mice. Note the NF- κ B nuclear staining in the hepatocytes and bile ducts of knockout mice and the decreased staining in ibuprofen-treated mice. All scale bars, 50 μ m. Knockout (KO) = *Mdr2*^{-/-}; WT, wild type. Arrows point to examples of: Kupffer cells (red arrows), NF- κ B-positive hepatocytes (yellow arrows) and bile epithelium (turquoise arrows).

Hepatocyte dysplasia has two major features, architectural disorganization and cytological atypia. Histological analysis revealed that anisocytosis was a prominent feature of both knockout and double mutant mice; dysplasia was evident at 4 months and increased in severity with age in both animal strains. We could not detect any difference in the extent or degree of dysplasia between knockout and double mutant mice at any age group (Fig. 3c). In an attempt to quantify the premalignant changes we analysed hepatocyte ploidy in both strains in comparison to age-matched wild-type mice. Isolated nuclei from livers were stained with propidium-iodide and subjected to fluorescence-activated cell sorting (FACS) analysis. Both groups had a high degree of hyperploidy, and we noted no significant difference between the cell-cycle profiles of the knockout and double mutant mice at 4 or 7 months (Fig. 3b). The finding that knockout and double mutant mice display comparable degrees of proliferation, hyperploidy and dysplasia implies that NF- κ B is not required for early neoplastic events.

Whereas NF- κ B inhibition had no effect at the tumour initiation phases, magnetic resonance imaging (MRI) and histological analysis of older mice clearly discerned the knockout and Dox-treated mice from the untreated (NF- κ B deficient) double mutant mice. At 10 months, 60% and 78% of knockout and Dox-treated double mutant mice, respectively, had liver tumours, compared with only 10% of untreated double mutant mice ($P < 0.01$, Fig. 4). Thus, ablation of NF- κ B activity in the hepatocytes led to a dramatic decrease in tumour progression. A similar tumour suppression effect in the NF- κ B-deficient animals was noted when comparing the mean number of tumours per animal for the two mouse groups.

Histological analysis of the livers confirmed the MRI data (data not shown).

At which stage of tumour development does NF- κ B inhibition exert its anti-tumorigenic effect? By 7 months no difference in early tumour development was noticed between knockout and double mutant mice: neither group had detectable tumours; only dysplastic parenchyma. We therefore switched off the transgene at 7 months, allowing NF- κ B activation in the hepatocytes. Tumour development was monitored using MRI for 3 months. Surprisingly, 7 of 9 mice fed on Dox (NF- κ B proficient) developed MRI-detectable tumours within this 3-month period. These were indistinguishable in size (Fig. 4b) and histological appearance (data not shown) from knockout tumours at a comparable age. Hence, it seems that whereas NF- κ B is dispensable for the early premalignant phase, (that is, tumour initiation) it is essential for subsequent tumour promotion.

Finally, how does NF- κ B support tumour promotion? NF- κ B has a major anti-apoptotic effect in the mouse liver during development¹⁶ and is necessary to protect mature hepatocytes against immune attack¹¹ and genotoxic stress¹⁷. Compromising the anti-apoptotic function of NF- κ B could contribute to the tumour-suppressing effect of the I κ B-super-repressor in the double mutant mice. To study this possibility, we assessed apoptosis using immunostaining for activated caspase-3. Knockout livers at 7 months (in which the majority of hepatocytes are dysplastic; Fig. 3c) showed more apoptosis than wild-type livers—probably due to the toxic effects of inflammation (Fig. 5a). However, blocking hepatocyte NF- κ B induced a further threefold increase in hepatocyte apoptosis.

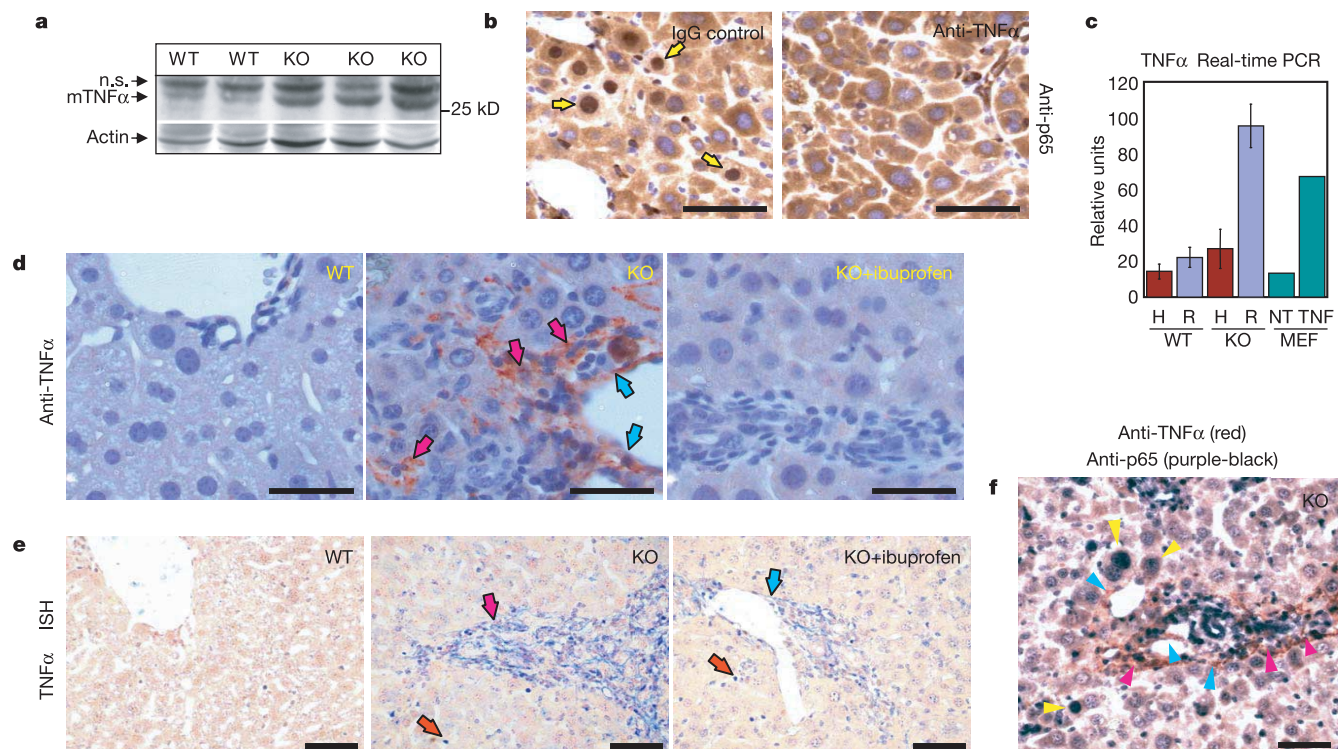


Figure 2 NF- κ B activation in *Mdr2*-knockout mice is induced by paracrine TNF α . **a**, Western blot analysis of TNF α in knockout versus wild-type liver samples. Anti-TNF α antibodies detect only the transmembrane TNF α (mTNF α ; 26 kD), the expression of which is enhanced in the knockout samples. n.s., non specific band. **b**, Anti-p65 immunostaining of knockout livers from 7-month-old mice treated with anti-TNF α neutralizing antibodies or control IgG as indicated. Quantitative analysis revealed a fivefold reduction in the number of p65-positive hepatocyte nuclei. **c**, Real-time PCR analysis of TNF α in hepatocyte (H) and non-hepatocyte (R) cellular fractions of knockout and wild-type livers. TNF α -treated (TNF) and non-treated (NT) mouse embryo fibroblasts

(MEF) serve as controls. Values given are TNF α mRNA levels normalized to a ribosomal standard (mean $\pm$ s.e.m.). **d**, Immunostaining for TNF α ; note the marked endothelial and mononuclear cell staining in knockout mice. **e**, *In situ* mRNA hybridization (ISH) for TNF α ; note the abundant signals in endothelial, mononuclear and Kupffer cells, particularly in untreated knockout samples. **f**, Double immunostaining of knockout liver sections for TNF α (AEC, red precipitate) and p65 (NBT/BCIP, purple-black precipitate). All scale bars, 50 μ m. Arrows: red, Kupffer cells; yellow, hepatocytes; turquoise, endothelium; purple, inflammatory cells.

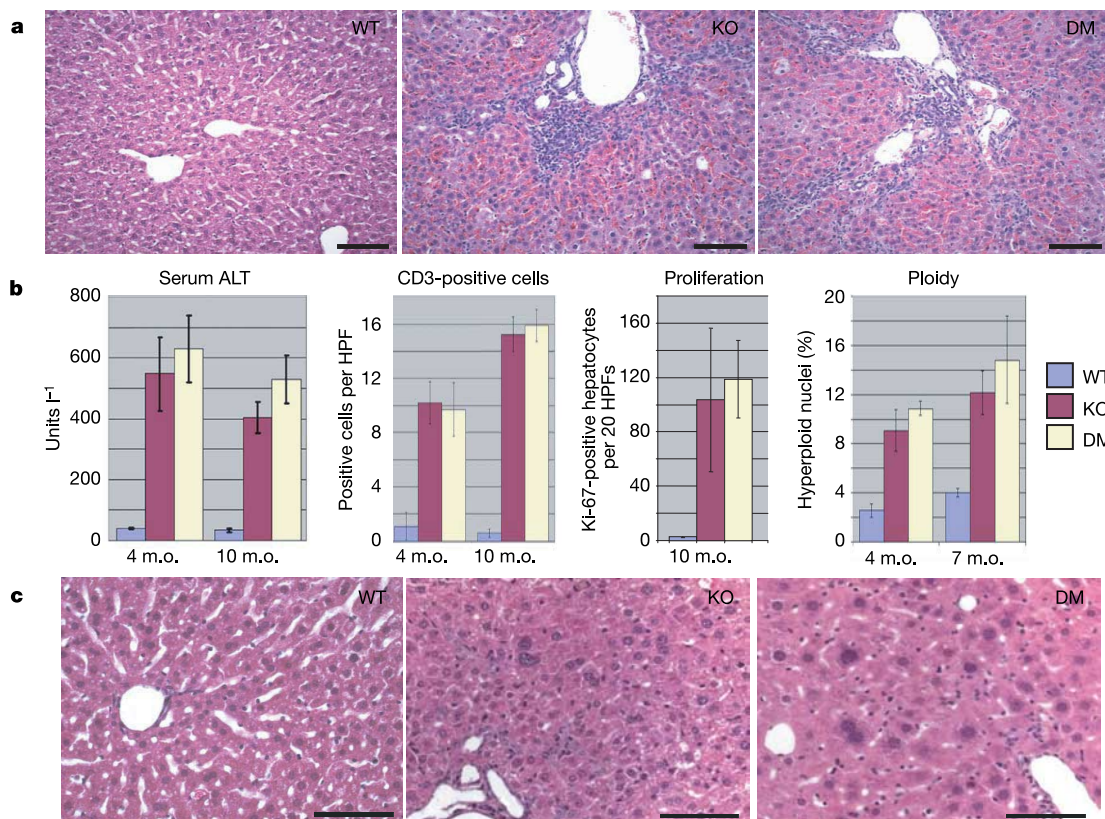


Figure 3 Inhibition of hepatocyte NF- κ B activity influences neither hepatitis nor development of the premalignant features in *Mdr2*-knockout mice. **a**, H&E-stained sections of 4-month-old wild-type, knockout and double mutant mice. **b**, Serum ALT levels, average number of CD3-positive cells per one high power field (HPF), number of Ki-67-positive hepatocytes per 20 HPFs and percentage of hyperploid nuclei. All

bar-graphs indicate mean $\pm$ s.e.m. **c**, H&E-stained sections from livers of 7-month-old mice of the indicated genotypes, showing comparable dysplasia evident by architectural disorganization and prominent nuclear pleomorphism in both knockouts and double mutants, in contrast to the normal nuclei in wild types. All scale bars, 100 μ m. m.o., months old; DM, double mutant.

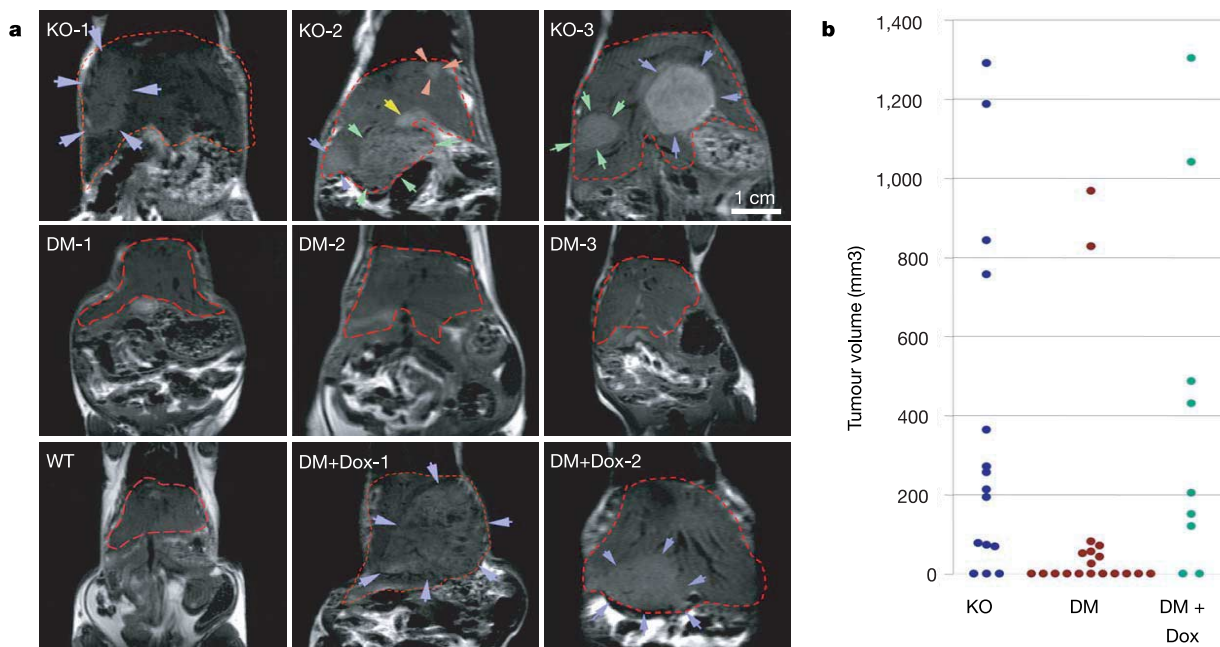


Figure 4 NF- κ B is dispensable for early stages of HCC development, but is critical for tumour promotion. **a**, MRIs from the indicated genotypes were taken at 10 months of age; each image shown is from a different animal. Red dashed lines outline the liver circumference. The coloured arrowheads indicate the tumours; different colours indicate individual tumours. **b**, Liver tumour volume was calculated by analysing all coronal and

axial images of each animal; nodules smaller than 100 mm³ were often found to represent dysplastic nodules rather than HCC in histological analysis and were therefore not scored as tumours. DM + Dox = mice fed on Dox between 7–10 months of age. Both knockout and double mutant + Dox groups differ significantly ($P < 0.01$, Mann–Whitney analysis) from the double mutant group, but not from each other.

Notably, a similar effect was achieved by anti-TNF α treatment. Taken together, our results suggest that NF- κ B inhibition exerts its tumour-suppressive effect by promoting apoptosis of transformed hepatocytes.

NF- κ B regulates multiple anti-apoptotic genes, some of which could be relevant to hepatocyte transformation. To identify relevant target genes and corroborate the pro-tumorigenic role of the TNF α -NF- κ B activation axis, expression of candidate genes was determined by semiquantitative PCR with reverse transcription (RT-PCR) and western blot analyses. Whereas some examined

targets (for example, XIAP) were similarly expressed in NF- κ B-proficient and -deficient livers, A1/Bfl1, c-IAP1 and GADD45 β were significantly enhanced in the knockout livers (Supplementary Fig. 3a, b). Concordantly, anti-TNF α treatment prevented the induction of GADD45 β and A1/Bfl1 in knockout mice (Fig. 5b).

It is conceivable that the inflammatory process in the liver contributes to tumorigenesis through two arms that complement one other (Fig. 5d). One arm would be responsible for inducing hepatocyte proliferation in knockout livers. Accelerated DNA replication is by itself error prone and a ground for further pro-tumorigenic alterations by genotoxic agents¹⁸, many of which are specifically activated in the liver (for example, aflatoxin)¹⁹. In addition, tumour promotion requires a second arm providing apoptosis protection through NF- κ B activation. Our data indicate that TNF α is a major driver of the anti-apoptotic arm, necessary for activation of NF- κ B-dependent anti-apoptotic genes, such as A1/Bfl1, c-IAP1 and GADD45 β . Although NF- κ B is activated early in the course of the disease, it is dispensable for a prolonged period during which premalignant foci accumulate in the inflamed liver, but is critical at the stage during which these foci turn malignant. Perhaps acquisition of oncogenic mutations during the tumour promotion phase renders premalignant hepatocytes particularly sensitive to apoptotic factors, which must be neutralized through NF- κ B activity. Putative apoptosis-regulating factors at this vulnerable tumour development phase are p53 and cJun (ref. 20), TNF receptor family members²¹ and cJun amino-terminal kinase (JNK)²². JNK activation and augmented cJun expression were observed in knockout hepatocytes at 7 months (Fig. 5c and Supplementary Fig. 3a), around the peak period of premalignant changes (Fig. 3). Similarly to NF- κ B, these were dependent on continuous TNF α signalling, as anti-TNF α treatment of knockout mice suppressed liver JNK activation and the expression of hepatocyte cJun (Fig. 5b, c). JNK activation is subject to NF- κ B inhibition via the NF- κ B target GADD45 β (refs 23, 24). NF- κ B inhibition in knockout hepatocytes could therefore enhance JNK-mediated apoptosis through inhibition of GADD45 β (Fig. 5b), and promote caspase-mediated hepatocyte apoptosis through suppression of multiple NF- κ B-induced anti-apoptotic genes (Supplementary Fig. 3). In parallel, cJun inhibition by anti-TNF α treatment (Fig. 5c) could promote p53-mediated apoptosis²⁰. Thus, by suppressing multiple anti-apoptotic pathways (Fig. 5d), anti-TNF α treatment may serve as a potential mode of HCC prophylaxis in chronic hepatitis patients.

Our conclusions are supported by recent studies in a colon inflammation model, in which colonic I κ B kinase β (IKK β) inactivation results in epithelial apoptosis during tumour promotion²⁵. Whereas inflammation releases many growth factors and cytokines^{1,26}, our data suggest that TNF α has a central role in activating NF- κ B and protecting transformed hepatocytes against apoptosis. It has been proposed that anticancer research might be more effective if aimed at eradicating the cause or the signalling context of abnormality rather than just treating the end result²⁷. Whereas eradicating the primary cause, for example chronic hepatitis, is currently unattainable, disrupting the signalling context of the evolving tumour may be a more realistic objective. Our finding that short-term ablation of epithelial NF- κ B activation is sufficient to induce premalignant cell apoptosis and halt tumour progression supports this conjecture. Intermittent suppression of a major signalling factor, such as NF- κ B or TNF α , could thus be a tool to protract the premalignant phase and inhibit tumour progression in chronic inflammatory diseases with high cancer risk. □

Methods

The following details can be found in the Supplementary Information: reagents, transgenic mouse construction and genotyping, semiquantitative RT-PCR, real-time PCR, *in situ* mRNA hybridization, liver cell fractionation, western blot and hepatocyte ploidy analyses.

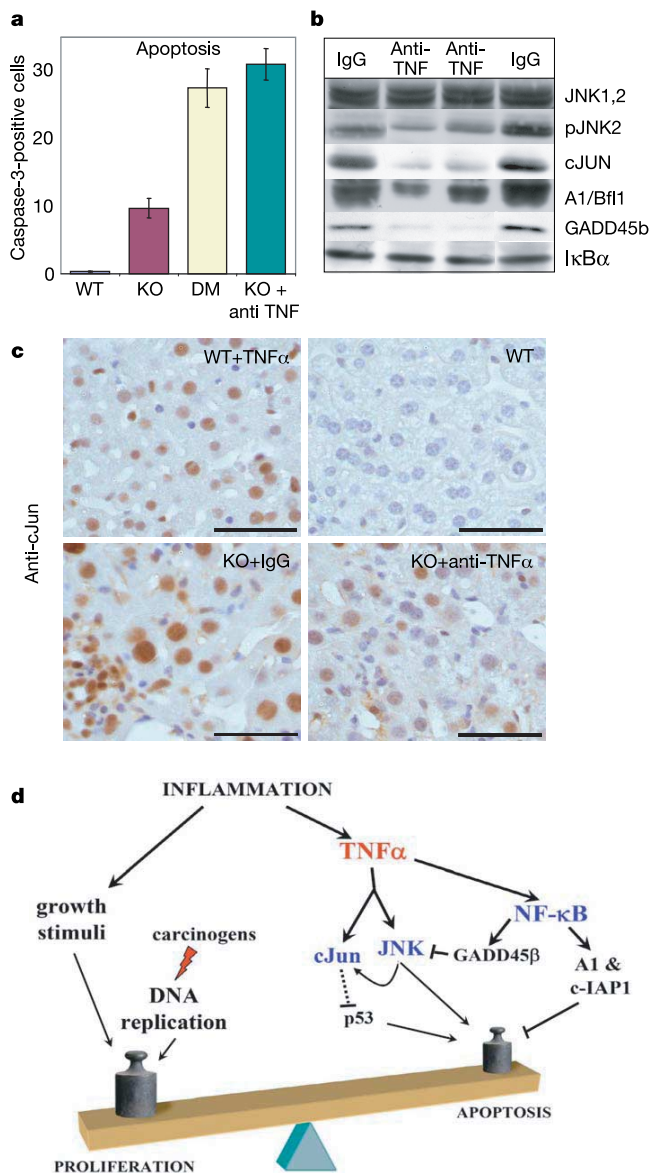


Figure 5 NF- κ B inhibition through the I κ B-super-repressor or anti-TNF α treatment results in suppression of anti-apoptotic factors. **a**, Caspase activity was quantified by counting the number of hepatocytes stained with antibodies against activated caspase-3 in 10 high-power fields (mean $\pm$ s.e.m.). **b**, Liver protein extracts from knockout animals treated for 3 days with anti-TNF α antibodies (anti-TNF), or IgG were subjected to western blot analysis with the indicated antibodies. **c**, cJun immunostaining of wild-type and knockout liver sections from IgG- and anti-TNF α -treated mice. TNF α -treated wild-type liver is a positive control for cJun staining. All scale bars, 50 μ m. **d**, A proposed model for inflammation-orchestrated signalling at the premalignant phase of HCC development: TNF α , a major mediator of the inflammatory process, tilts the balance in favour of hepatocyte proliferation. Carcinogenic metabolites would push tumour progression further towards malignancy.

Animal studies and tissue preparation

All animal experiments were performed in accordance with the guidelines of the institutional committee for the use of animals for research. Mice were killed by a lethal dose of anaesthesia. *In vivo* luciferase imaging and doxycycline treatment were performed as previously described¹¹. Animals were injected intraperitoneally (i.p.) with 20 ng human recombinant TNF α (Roche) and killed after 1 h for liver p65 immunostaining and RNA and protein extraction. For anti-TNF α treatment, mice were injected with 22 μ g of goat anti-mouse TNF α neutralizing antibodies i.p. (R&D Systems; AF-410-NA) for 3 consecutive days, or with control goat IgG. All mice were injected i.p. with BrdU (Amersham; 100 μ l per 10 g body weight) 3 and 24 h before they were killed. In some animals, liver samples were removed and snap-frozen for protein and RNA analyses. The animal was then perfused through the left ventricle with 10 ml of cold heparinized PBS, followed with 25 ml of 4% buffered formalin. Livers were removed, weighed, photographed and fixed in formalin overnight. The next day the entire liver was submitted for paraffin embedding in three to five cassettes. 5- μ m sections were stained with H&E and evaluated by a pathologist to whom the genetic makeup or treatment group were not known.

MRI analysis

MRI was performed on a horizontal 4.7T Biospec spectrometer (Bruker Medical), using a birdcage coil. Mice were anaesthetized (30 mg kg⁻¹ pentobarbital, i.p.) and placed supine with the liver located at the centre of the coil. Liver and tumour volumes were determined from multi-slice coronal and axial T1 weighted fast spin echo images covering the entire liver both coronally and axially (repetition time, 400 ms; echo time, 17 ms; slice thickness, 1 mm; field of view, 5 cm (coronal) and 3.4 cm (axial) using a 256 $\times$ 256 matrix). To estimate tumour and liver volumes, the liver and tumour boundaries visualized in each slice were outlined by using image processing software (NIH image), by an observer to whom the genetic makeup was not known. The number of pixels (for both liver and tumour) were converted to an area by multiplying by the factor ((field of view)² $\times$ (matrix)²).

Immunohistochemistry

For p65 immunostains, 5- μ m sections, cut on the same day, were de-waxed and hydrated through graded ethanols, cooked in 25 mM citrate buffer pH 6.0 in a pressure cooker at 115 °C for 3 min (decloaking chamber, Biocare Medical), transferred into boiling deionized water and left to cool down for 20 min. After 5 min treatment in 3% H₂O₂, slides were incubated with rabbit polyclonal p65 antibodies diluted 1:100 in CAS-Block (Zymed) for 3 h at room temperature, washed three times with Optimax (Biogenex), incubated for 30 min with anti rabbit Envision⁺ (DAKO) and developed with DAB for 15 min. Antigen retrieval for TNF α and p65 double immunostaining was performed in 100 mM glycine buffer pH 9.0. After 5 min treatment in 3% H₂O₂, slides were incubated with goat polyclonal anti-TNF α antibodies diluted 1:50 in CAS-Block for 1 h at room temperature, washed in Optimax, incubated with (1:100) rabbit anti-goat antibodies (Jackson Laboratories) for 30 min, washed again with Optimax and incubated with anti-rabbit Envision⁺ (DAKO) and developed with AEC for 15 min at 37 °C. Following this, slides were reboiled for 7 min in 100 mM glycine buffer in a microwave oven and cooled down to room temperature. Slides were incubated with rabbit polyclonal anti-p65 antibodies diluted 1:100 in CAS-Block for 3 h at room temperature, washed three times with Optimax, post-fixed for 5 min in 4% formaldehyde in Tris buffered saline, washed with Optimax, incubated for 30 min with alkaline phosphatase-conjugated polymer anti-rabbit IgG (Zymed) and developed with BCIP/NBT. Slides were counterstained with nuclear fast red. Antigen retrieval for CD3, cJun, BrdU, activated caspase-3, luciferase, MPO and Ki-67 were performed in 25 mM citrate buffer pH 6.0. Antigen retrieval for TNF α was performed in 100 mM glycine buffer pH 9.0. Detailed protocols for immunostaining with each antibody are available on request.

Received 14 June; accepted 10 August 2004; doi:10.1038/nature02924.
Published online 25 August 2004.

- Coussens, L. M. & Werb, Z. Inflammation and cancer. *Nature* **420**, 860–867 (2002).
- Balmain, A. & Harris, C. C. Carcinogenesis in mouse and human cells: parallels and paradoxes. *Carcinogenesis* **21**, 371–377 (2000).
- Li, Q. & Verma, I. M. NF- κ B regulation in the immune system. *Nature Rev. Immunol.* **2**, 725–734 (2002).
- Mayo, M. W. & Baldwin, A. S. The transcription factor NF- κ B: control of oncogenesis and cancer therapy resistance. *Biochim. Biophys. Acta* **1470**, M55–M62 (2000).
- Lin, A. & Karin, M. NF- κ B in cancer: a marked target. *Semin. Cancer Biol.* **13**, 107–114 (2003).
- Mauad, T. H. *et al.* Mice with homozygous disruption of the mdrl2 P-glycoprotein gene. A novel animal model for studies of nonsuppurative inflammatory cholangitis and hepatocarcinogenesis. *Am. J. Pathol.* **145**, 1237–1245 (1994).
- Nakamoto, Y., Guidotti, L. G., Kuhlén, C. V., Fowler, P. & Chisari, F. V. Immune pathogenesis of hepatocellular carcinoma. *J. Exp. Med.* **188**, 341–350 (1998).
- Block, T. M., Mehta, A. S., Fimmel, C. J. & Jordan, R. Molecular viral oncology of hepatocellular carcinoma. *Oncogene* **22**, 5093–5107 (2003).
- De Vree, J. M. *et al.* Correction of liver disease by hepatocyte transplantation in a mouse model of progressive familial intrahepatic cholestasis. *Gastroenterology* **119**, 1720–1730 (2000).
- Staudt, L. M. Molecular diagnosis of the hematologic cancers. *N. Engl. J. Med.* **348**, 1777–1785 (2003).
- Lavon, I. *et al.* High susceptibility to bacterial infection, but no liver dysfunction, in mice compromised for hepatocyte NF- κ B activation. *Nature Med.* **6**, 573–577 (2000).
- Gupta, S. Hepatic polyploidy and liver growth control. *Semin. Cancer Biol.* **10**, 161–171 (2000).
- Bruix, J., Boix, L., Sala, M. & Llovet, J. M. Focus on hepatocellular carcinoma. *Cancer Cell* **5**, 215–219 (2004).
- Chaisson, M. L., Brooling, J. T., Ladiges, W., Tsai, S. & Fausto, N. Hepatocyte-specific inhibition of NF- κ B leads to apoptosis after TNF treatment, but not after partial hepatectomy. *J. Clin. Invest.* **110**, 193–202 (2002).

- Maeda, S. *et al.* IKK β is required for prevention of apoptosis mediated by cell-bound but not by circulating TNF α . *Immunity* **19**, 725–737 (2003).
- Beg, A. A., Sha, W. C., Bronson, R. T., Ghosh, S. & Baltimore, D. Embryonic lethality and liver degeneration in mice lacking the RelA component of NF- κ B. *Nature* **376**, 167–170 (1995).
- Lavon, I. *et al.* Nuclear factor- κ B protects the liver against genotoxic stress and functions independently of p53. *Cancer Res.* **63**, 25–30 (2003).
- Friedberg, E. C., McDaniel, L. D. & Schultz, R. A. The role of endogenous and exogenous DNA damage and mutagenesis. *Curr. Opin. Genet. Dev.* **14**, 5–10 (2004).
- Wang, J. S. & Groopman, J. D. DNA damage by mycotoxins. *Mutat. Res.* **424**, 167–181 (1999).
- Eferl, R. *et al.* Liver tumor development. c-Jun antagonizes the proapoptotic activity of p53. *Cell* **112**, 181–192 (2003).
- Ehrhardt, H. *et al.* TRAIL induced survival and proliferation in cancer cells resistant towards TRAIL-induced apoptosis mediated by NF- κ B. *Oncogene* **22**, 3842–3852 (2003).
- Kennedy, N. J. & Davis, R. J. Role of JNK in tumor development. *Cell Cycle* **2**, 199–201 (2003).
- Tang, G. *et al.* Inhibition of JNK activation through NF- κ B target genes. *Nature* **414**, 313–317 (2001).
- Papa, S. *et al.* Gadd45 β mediates the NF- κ B suppression of JNK signalling by targeting MKK7/JNK2. *Nature Cell Biol.* **6**, 146–153 (2004).
- Greten, F. R. *et al.* Ikk β links inflammation and tumorigenesis in a mouse model of colitis-associated cancer. *Cell* **118**, 285–296 (2004).
- Wilson, J. & Balkwill, F. The role of cytokines in the epithelial cancer microenvironment. *Semin. Cancer Biol.* **12**, 113–120 (2002).
- Radisky, D., Hagios, C. & Bissell, M. J. Tumors are unique organs defined by abnormal signaling and context. *Semin. Cancer Biol.* **11**, 87–95 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements *Mdr2*-knockout mice were a gift from R. P. Oude Elferink and the TA^{LAP1} mice were received from H. Bujard. We are grateful to N. Berger, T. Golub and N. Kidess-Bassir for technical assistance, to A. Hatzubai, O. Mandilboim, N. Lieberman, G. Kojekaro, M. Davis and H. Harel for helping with MRI, FACS, mRNA and protein analysis. We thank K. Meir and M. Oren for discussions and a critical reading of the manuscript. This research was supported by grants from the Israel Science Foundation funded by the Israel Academy for Sciences and Humanities (Center of Excellence Program), Prostate Cancer Foundation Israel—Center of Excellence, German-Israeli Foundation for Scientific Research and Development (GIF, in collaboration with H. Bujard), a grant in memory of H. and F. Brody from H. M. Krueger as trustee of a charitable trust and the Israel Cancer Research Foundation (ICRF). I.S. is supported by the Lady Davis Fellowship Trust.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.P. (peli@hadassah.org.il) and Y.B.-N. (yinon@cc.huji.ac.il).

Mrf4 determines skeletal muscle identity in *Myf5:Myod* double-mutant mice

Lina Kassas-Duchosoy¹, Barbara Gayraud-Morel^{1*}, Danielle Gomes¹,
Didier Rocancourt², Margaret Buckingham², Vasily Shinin^{1*}
& Shahragim Tajbakhsh¹

¹Stem Cells and Development, ²Molecular Genetics of Development,
Department of Developmental Biology, CNRS URA 2578, 25 rue du Dr Roux,
75724 Paris Cedex 15, France

* These authors contributed equally to the work

In vertebrates, skeletal muscle is a model for the acquisition of cell fate from stem cells¹. Two determination factors of the basic helix–loop–helix myogenic regulatory factor (MRF) family, *Myf5* and *Myod*, are thought to direct this transition because double-mutant mice totally lack skeletal muscle fibres and myoblasts^{2–4}. In the absence of these factors, progenitor cells remain multipotent and can change their fate^{5,6}. Gene targeting studies have revealed hierarchical relationships between these and the other MRF genes, *Mrf4* and *myogenin*, where the latter are regarded as differentiation genes⁷. Here we show, using an allelic series of three *Myf5* mutants that differentially affect the expression of the genetically linked *Mrf4* gene, that skeletal muscle is present in the

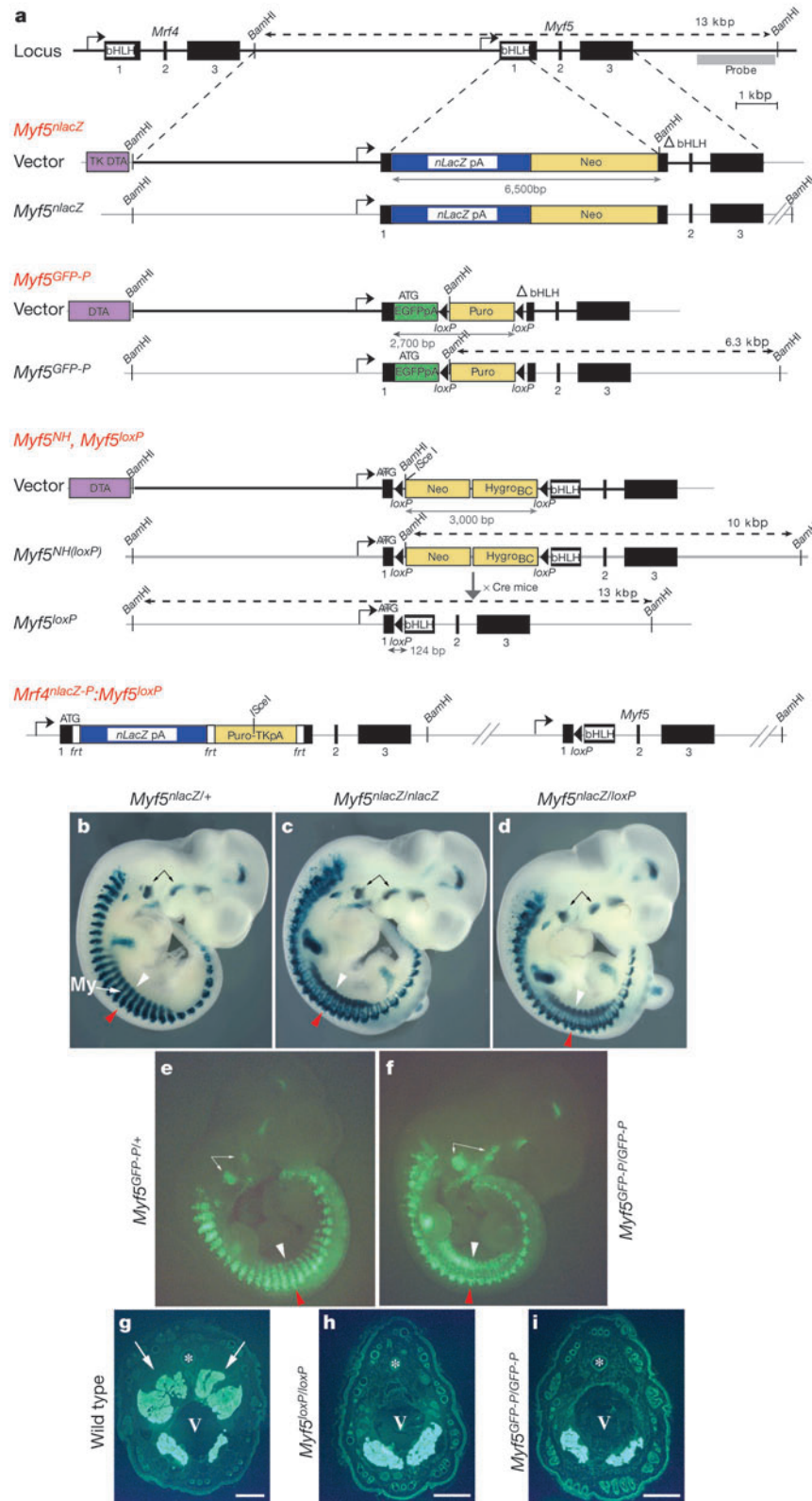


Figure 1 Allelic series of *Myf5* mutants. **a**, Both *Myf5^{nlacZ}* and *Myf5^{GFP-P}* contain a deletion of the N terminus and bHLH domain. *Mrf4* was targeted with an *nlacZ* reporter gene (Methods). **b–i**, Early myotome is perturbed in *Myf5*-null embryos. Myotome (My) is present in somites of heterozygous E10.5 embryos (**b**, **e**) and is perturbed in null mutants (**c**, **d**, **f**). Null mutants display mislocalized MPCs along the epaxial (dorsal, red arrowhead) and hypaxial (ventral, white arrowhead) somites. Arrows indicate branchial arch muscles

(**b–f**). Tails of adult wild-type (**g**), *Myf5^{loxP/loxP}* (**h**) and *Myf5^{GFP-P/GFP-P}* (**i**) animals were sectioned on a cryostat and stained with anti-MyHC. Note that dorsal tail muscles are lacking in the mutants (arrows). Asterisk, dorsal vein; kbp, kilobase pairs; TK DTA, diphtheria toxin gene with thymidine kinase promoter; pA, SV40 polyA; Puro-TKpA, Puromycin-thymidine kinase fusion with TK promoter and bovine polyA; V, vertebra. Scale bar, 300 μ m.

new *Myf5:Myod* double-null mice only when *Mrf4* expression is not compromised. This finding contradicts the widely held view that myogenic identity is conferred solely by *Myf5* and *Myod*, and identifies *Mrf4* as a determination gene. We revise the epistatic relationship of the MRFs, in which both *Myf5* and *Mrf4* act upstream of *Myod* to direct embryonic multipotent cells into the myogenic lineage.

Segmented structures called somites provide muscle progenitor cells (MPCs) for the skeletal muscles of the trunk, limbs and tail, whereas head muscles are derived predominantly from paraxial head and prechordal mesoderm⁷. As somites mature in an anterior–posterior gradient, they transiently retain a dorsal dermomyotome epithelium, which is the source of resident skeletal muscle cells constituting the myotome and of migratory MPCs¹. *Myf5* and *Myod* are thought to initiate muscle identity in MPCs.

Our analyses of several *Myf5* mutants generated by homologous recombination indicated that *Mrf4* is expressed in MPCs, and its expression was differentially affected in an allelic series comprising the previously described *Myf5^{nlacZ}* allele⁵ and two newly generated *Myf5^{GFP-P}* and *Myf5^{loxP}* alleles (Fig. 1a). The null status of these

alleles was examined *in vitro* and *in vivo*. *Myf5^{nlacZ}* and *Myf5^{GFP-P}* both lack the basic helix–loop–helix (bHLH) domain required for DNA binding and dimerization and are therefore null⁸. The *Myf5^{loxP}* allele is also null because it lacks the amino-terminal domain required for transactivation (Supplementary Figs 1, 2). As with all *Myf5*-null embryos^{3,5,9,10}, these mutants lack an early myotome until at least embryonic day 10 (E10; Fig. 1c, d, f; see below), compared with normal embryos (Fig. 1b, e). Another characteristic of *Myf5*-null embryos is a mislocalization of MPCs along the epaxial–most (dorsal) and hypaxial–most (ventral) dermomyotome lips, particularly evident in interlimb somites (Fig. 1c, d, f). In addition, *Myf5*-null mutants lack epaxial (dorsal) muscles^{7,11} (Fig. 1g–i; data not shown).

To determine which MRF rescues the myotome in *Myf5* mutants, we examined *Mrf4* and *Myod* expression by whole-mount *in situ* hybridization (WM-ISH). *Mrf4* expression initiates from E9.5 in wild-type embryos, and by E9.75 *Mrf4* transcripts accumulate in the epaxial and hypaxial somite. Epaxial somite expression marks predominantly differentiating cells in the myotome that were previously programmed by *Myf5*. By contrast, hypaxial *Mrf4*

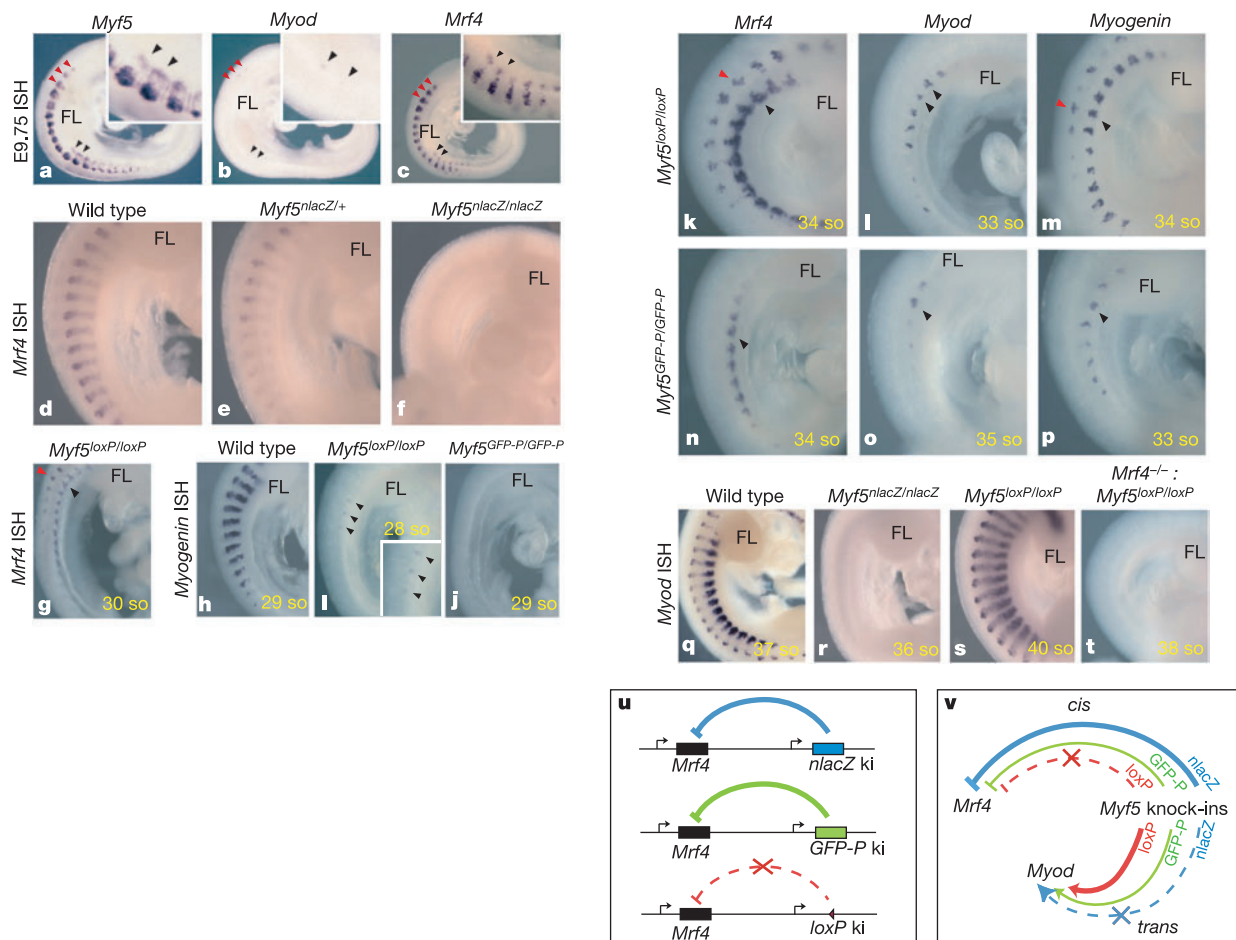


Figure 2 *Myod* expression depends on *Mrf4* in *Myf5* mutants. **a–c**, WM-ISH of E9.75 (27–28 somites) wild-type embryos with *Myf5* (**a**), *Myod* (**b**) and *Mrf4* (**c**) riboprobes. *Myf5* and *Mrf4* are expressed before *Myod* in the epaxial and hypaxial somite. **d–f**, *Mrf4* transcripts in somites of E10.5 wild-type (**d**), *Myf5^{nlacZ/+}* (**e**) and *Myf5^{nlacZ/nlacZ}* (**f**) embryos. **g**, *Mrf4* ISH on E10 *Myf5^{loxP/loxP}* embryo. **h–j**, The *myogenin* gene is not expressed at E9.75 in *Myf5^{loxP/loxP}* and *Myf5^{GFP-P/GFP-P}* compared with wild type. Note start of hypaxial expression in *Myf5^{loxP/loxP}* (arrowheads). **k–p**, *Mrf4* (**k**, **n**), *Myod* (**l**, **o**) and *myogenin* (**m**, **p**) ISHs on E10.5 *Myf5^{loxP/loxP}* (**k–m**) and *Myf5^{GFP-P/GFP-P}* (**n–p**) embryos. Note that *Myod* and *myogenin* gene expression appear after *Mrf4*; the central somite lacks

signal owing to misplaced MPCs (see Fig. 1d, f). *Mrf4* and *Myod* expression in interlimb somites begins in the hypaxial somite, unlike *Myf5*, and later in the epaxial somite. **q–t**, *Myod* transcripts are absent at E10.5 in *Myf5^{nlacZ/nlacZ}* (**r**) and *Mrf4^{-/-}* (**t**) mutants compared with controls (**q**, **s**). **u–v**, Perturbations of *Mrf4* in cis (**u**) correlate with perturbations of *Myod* in trans (**v**). The curved line thickness in **u** and **v** reflects the relative intensity of inhibition or activation, such that the greatest reduction in *Mrf4* expression (**u**, **v**) is observed with *Myf5^{nlacZ/nlacZ}*, which correlates with the greatest delay in *Myod* expression (**v**). FL, forelimb; ki, *Myf5* knock-ins; so, somite; red arrowheads, epaxial (dorsal); black arrowheads, hypaxial (ventral).

expression was observed concomitantly with *Myf5* in the dermomyotome¹², before myotome formation and before *Myod* expression (Figs 2a–c, 3g, h; data not shown), suggesting that *Mrf4* might have a function in MPCs. Strikingly, examination of *Myf5* mutants demonstrated that *Mrf4* expression was directly dependent on the nature and size of the knock-in generated at the *Myf5* locus, suggesting that these perturbations are *in cis* (Fig. 2u). In the most severe case, *Mrf4* expression was not detected in *Myf5*^{nlacZ/nlacZ} embryos at E10.5, compared with robust expression in wild-type embryos (Fig. 2d, f). Consistent with this, *Mrf4* expression in *Myf5*^{nlacZ/+} embryos was about half of that observed in the wild type (Fig. 2d–f). These observations suggest that *Myf5*^{nlacZ/nlacZ} mutants correspond to a loss of function of both *Myf5* and *Mrf4* at these early stages. By contrast, *Mrf4* was expressed in both *Myf5*^{loxP/loxP} and *Myf5*^{GFP-P/GFP-P} embryos (Fig. 2g, k, n). The onset of *Mrf4* expression was not appreciably affected in the hypaxial somite of *Myf5*^{loxP/loxP} embryos at E10 (Fig. 2c, g), whereas

Mrf4 was not detected at this stage in *Myf5*^{GFP-P/GFP-P} embryos but was observed at later stages (Fig. 2n; data not shown). As in *Myf5*^{nlacZ/nlacZ} embryos, differentiated cells were lacking in these mutants at these early stages (Fig. 2i, j).

We showed previously that *Myod* expression in somites was delayed until after E10.5 in *Myf5*^{nlacZ/nlacZ} embryos, and proposed that *Myf5* acts upstream of *Myod*¹³ (Fig. 2q, r; *n* = 10). We now show that in *Myf5* mutants in which *Mrf4* expression is not compromised, *Myod* expression is detectable before E10.5 (Fig. 2l, o, s; *n* = 18). Furthermore, as in wild-type embryos, comparative WM-ISH analysis showed that *Mrf4* expression precedes that of *Myod* in these mutants (Fig. 2k, l, n, o; *n* = 12), suggesting that *Mrf4* as well as *Myf5* might participate in the activation of *Myod*. Similar results were obtained with a fourth *Myf5*-null allele (*Myf5*^{dCT/dCT}; data not shown). Given these findings, we would predict that the direct inactivation of both *Mrf4* and *Myf5* would phenocopy *Myf5*^{nlacZ/nlacZ} with respect to *Myod* expression. To test this, and to distinguish between effects *in cis* and *in trans*, we examined *Myod* expression in mutants lacking functional *Mrf4* as well as *Myf5* (Fig. 1a). *Myod* expression was not detected at E10.5 in *Myf5*^{nlacZ/nlacZ} or *Mrf4*^{nlacZ-P/nlacZ-P}:*Myf5*^{loxP/loxP} embryos compared with controls (Fig. 2q–t). Furthermore, *Myod* expression levels in *Mrf4*^{nlacZ-P/+}:*Myf5*^{loxP/+} mutants were about half those observed in the wild type (data not shown). In addition, the expression of *desmin* (which marks undifferentiated myoblasts as well as differentiated muscle), *myogenin*, and differentiation markers was lacking before E10.5 in *Mrf4*^{nlacZ-P/nlacZ-P}:*Myf5*^{loxP/loxP} and *Myf5*^{nlacZ/nlacZ} embryos (Fig. 3d–i; data not shown). In contrast, *desmin* and *myogenin* were expressed by this stage in *Myf5*^{GFP-P/GFP-P} and *Myf5*^{loxP/loxP} embryos (Figs 2m, p, 3b, c). Notably, *desmin* and *myogenin* expression were absent before E9.5 in *Myf5*^{loxP/loxP} and *Myf5*^{GFP-P/GFP-P} mutants (Fig. 2h–j; data not shown; *n* = 8). Their expression appeared after that of *Mrf4* transcript and protein, and before the expression of *Myod* (Figs 2g, k, l, n, o, 3a–c; *n* = 12, data not shown). A graded delay in *Myod* expression and myotome formation was therefore observed such that *Myf5*^{nlacZ/nlacZ} > *Myf5*^{GFP-P/GFP-P} > *Myf5*^{loxP/loxP}, and this was correlated with *Mrf4* expression. These results suggest that *Mrf4* can determine muscle identity and acts upstream of *Myod* (Fig. 2v). To test whether *Mrf4* can determine muscle identity autonomously, we examined *Myf5*:*Myod* double mutants.

Extensive analysis of *Myf5*^{nlacZ/nlacZ}:*Myod*^{-/-} mutants failed to demonstrate skeletal myogenesis with a number of muscle markers at all developmental stages⁴ (data not shown). However, robust skeletal muscle differentiation was observed at E12.5 in *Myf5*^{loxP/loxP}:*Myod*^{-/-} and *Myf5*^{GFP-P/GFP-P}:*Myod*^{-/-} embryos in which *Mrf4* expression was not fully compromised (Fig. 4a–d). Muscle differentiation was detected in both the epaxial and hypaxial regions (Fig. 4b, d, arrowheads), which was consistent with the observation that MPCs of both express *Mrf4* (Figs 2g, 3g, h). The desmin, myogenin and troponin T proteins were also observed in these differentiated cells (Fig. 4e–i; data not shown). Skeletal muscle was not detected in the limbs of these double mutants at this stage (Fig. 4d; data not shown), in keeping with previous observations that skeletal muscle formation is delayed until E13.5 in limbs of *Myod*^{-/-} embryos¹¹. Although some MyHC⁺ cells and skeletal muscle fibres were detected in the limbs at E14.5, fetal myogenesis was severely compromised so that the new *Myf5*:*Myod* double mutants were born essentially devoid of skeletal muscle. It is notable that skeletal myogenesis in the head was absent in the new double mutants (data not shown).

Previous work demonstrated that *Myf5*:*Myod* double mutants, and those carrying only one functional allele of *Myf5* or *Myod*, died postnatally². Our examination of the new *Myf5*:*Myod* mutants in which *Mrf4* was functional revealed that only the homozygous double mutant was lethal, whereas viable mice were generated from all other allelic combinations (data not shown). Taken together,

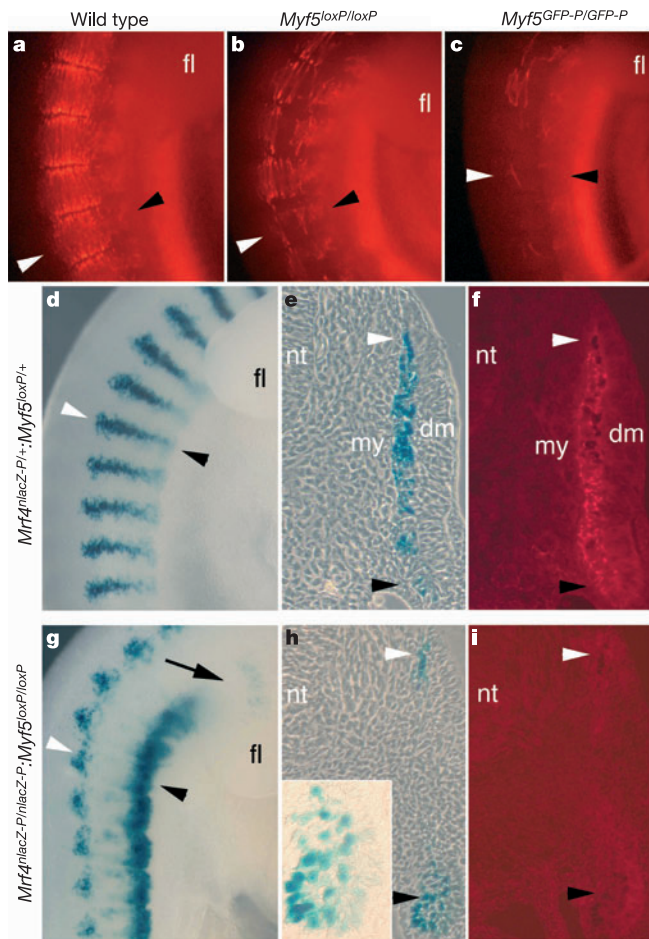


Figure 3 Myotome formation and desmin staining in *Myf5*-null and *Mrf4*:*Myf5* double-mutant embryos. **a–c**, Progressively less anti-desmin staining in whole mount of E10.5 wild-type (**a**), *Myf5*^{loxP/loxP} (**b**) and *Myf5*^{GFP-P/GFP-P} (**c**) embryos. Note mislocalized muscle precursors along the epaxial and hypaxial somite in the mutants. **d**, X-Gal staining of E10.5 *Mrf4*^{nlacZ-P/+}:*Myf5*^{loxP/+} embryo. **e**, Phase-contrast image of transverse cryostat section of **d**. **f**, Anti-desmin (red) staining of section in **e** showing correct myotome formation. **g**, X-Gal staining of E10.5 *Mrf4*^{nlacZ-P/nlacZ-P}:*Myf5*^{loxP/loxP} embryo showing MPCs mislocalized along the epaxial and hypaxial somite. Note that *Myf5* is not essential for *Mrf4* activation. **h**, Phase contrast image of section in **g** showing *Mrf4* expression in the dermomyotome epithelium (inset). **i**, absence of desmin corresponds to lack of myotome. dm, dermomyotome; fl, forelimb; my, myotome; nt, neural tube; white arrowheads, epaxial; black arrowheads, hypaxial; arrow, *Mrf4*⁺ cells in limb.

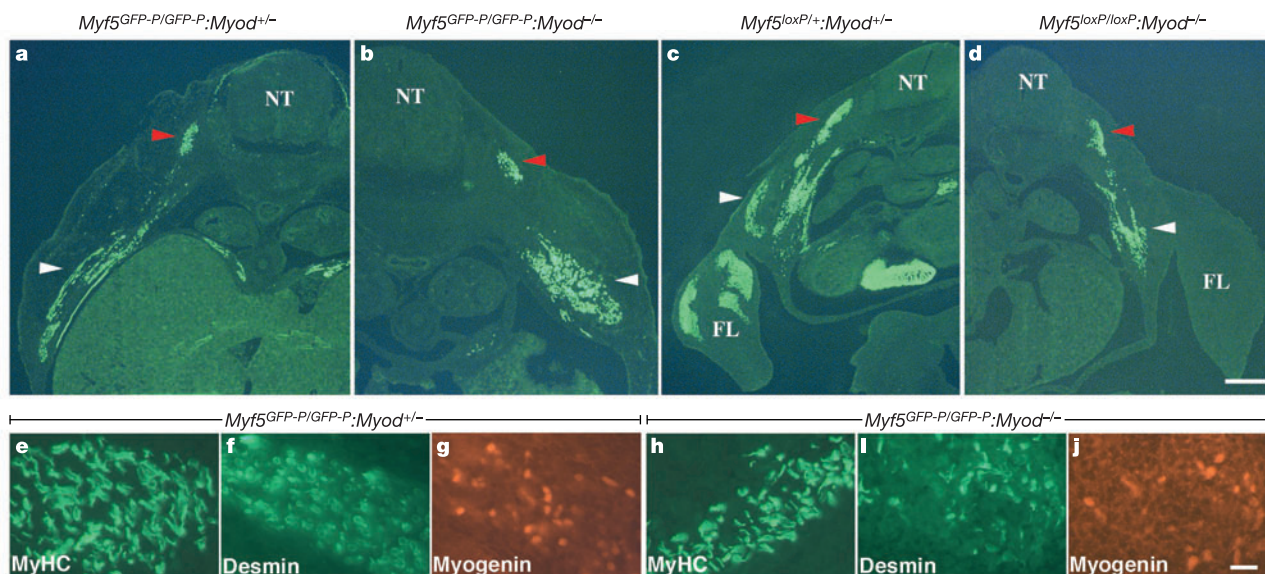


Figure 4 *Myf5:Myod* double mutants make muscle. **a–d**, Transverse cryostat sections of E12.5 control *Myf5^{GFP-P/GFP-P};Myod^{+/-}* (**a**), *Myf5^{loxP/+};Myod^{+/-}* (**c**), and double-null mutant *Myf5^{GFP-P/GFP-P};Myod^{-/-}* (**b**), *Myf5^{loxP/loxP};Myod^{-/-}* (**d**) embryos stained with anti-MyHC antibody. Note lack of limb muscles in **d** compared with **c** at this stage.

e–j, High magnification of anti-MyHC (**e, h**) anti-desmin (**f, i**) and anti-myogenin (**g, j**) antibody stainings of control (**e–g**) and double-null (**h–j**) embryos. Arrowheads indicate epaxial (red) and hypaxial (white) muscles. FL, forelimb; NT, neural tube. Scale bar, 300 μ m (**a–d**); 30 μ m (**e–j**).

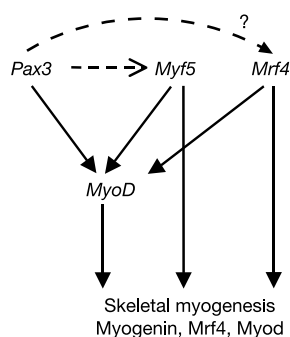


Figure 5 Genetic hierarchy governing the acquisition of skeletal muscle identity in the mouse. *Pax3*, *Myf5* and *Mrf4* act upstream of *MyoD* for skeletal myogenesis in the trunk. *Mrf4* can direct embryonic, but not fetal, skeletal muscle identity and differentiation in the absence of *Myf5* and *MyoD*. *MyoD* is initially activated by *Myf5* and *Mrf4*, and later through *Pax3* (ref. 13). *Mrf4* drives myogenesis in the embryonic trunk and limbs but not in the head or the fetus. *Myf5*, *MyoD* and *Mrf4* are determination genes. Arrows depict genetic relationships, and do not necessarily imply direct interactions.

these findings reinforce the notion that both the nature of the myogenic factor and the total MRF levels are crucial for ensuring postnatal viability.

We also examined whether myogenesis would be enhanced or deregulated in culture, as with *myogenin* null embryos in which differentiated fibres are made *in vitro* but not in the embryo^{14,15}. Monodisperse micromass cultures were made from all three double-mutant embryos at E11.5: *Myf5^{nlacZ/nlacZ};Myod^{-/-}*, *Myf5^{GFP-P/GFP-P};Myod^{-/-}* and *Myf5^{loxP/loxP};Myod^{-/-}*. Although skeletal myogenesis was readily observed with both *Myf5^{GFP-P/GFP-P};Myod^{-/-}* and *Myf5^{loxP/loxP};Myod^{-/-}* by using various markers, neither desmin⁺ cells nor skeletal muscle differentiation was detected in cultures of *Myf5^{nlacZ/nlacZ};Myod^{-/-}* double-mutant embryos (Supplementary Fig. 3a–e; data not shown; *n* = 7), which is consistent with the *in vivo* observations^{2–4}. Examination by time-lapse microscopy and imaging of *Myf5^{GFP-P/GFP-P};Myod^{-/-}* double mutants showed no detectable difference from controls in the ability of these cells to migrate on the dish or in their ability to

divide (Supplementary Fig. 3f–h).

Perturbations *in cis*, resulting from the insertion of foreign sequences, have been reported for several genes, including *Mrf4* (refs 16–18), where *Myf5* expression is affected. This locus is regulated by a complex array of proximal and distal regulatory elements^{19,20}. We now show that gene manipulations in *Myf5* can alter the regulated expression of *Mrf4*. We showed previously that *Pax3* and *Myf5* act upstream of *Myod* because body muscles are lacking in *Pax3^{Sp/Sp};Myf5^{nlacZ/nlacZ}* embryos¹³. Our present studies show that both *Mrf4* and *Myf5* act genetically upstream of *Myod* for embryonic skeletal myogenesis in the body. In the absence of these factors, *Myod* expression is delayed, probably until *Pax3*-mediated events activate this gene^{13,21}. Furthermore, our demonstration that *Myf5^{nlacZ/nlacZ}* mutants correspond to double *Myf5:Mrf4* gene inactivations at early stages suggests that, in *Pax3:Myf5^{loxP/loxP}* double mutants, which allow for *Mrf4* expression, *Myod* should drive skeletal myogenesis in the trunk (Fig. 5). Our preliminary observations suggest that this is so. Our finding that *Mrf4* does not direct skeletal muscle identity in the head points to a molecular heterogeneity in skeletal muscle stem cells at different anatomical locations. The recent finding that *MyoR:capsulin* double mutants lack a subset of head muscles also supports this notion²².

In a recently described *Myf5^{Δlox/Δlox};Myod^{-/-}* double mutant, skeletal muscle was not detected at birth³; it is possible that embryonic myogenesis is rescued in this mutant. Our findings also suggest that the insertion of a *myogenin* cDNA into the *Myf5* locus would permit *Mrf4* expression²³. Indeed, the muscle perturbations in *Myf5^{myg-ki/myg-ki};Myod^{-/-}* resemble those of the new *Myf5:Myod* double mutants described here. We also propose that *Mrf4* or *Myod* is necessary to activate the *myogenin* gene, because *Mrf4:Myod* double mutants phenocopy the *myogenin* null phenotype¹⁵.

Finally, our findings provide *in vivo* evidence consistent with previously unexplained molecular observations showing that the chromatin-remodelling domain of *Mrf4*, but not that of *myogenin*, can substitute for *Myod*²⁴. Those studies *in vitro*, and ours *in vivo* now support the notion that the formation of skeletal muscle depends on three determination factors which direct multipotent progenitors into the myogenic lineage. □

Methods

Gene targeting, generation of mutant mice, and genotyping

Embryonic stem cells (CK35, provided by C. Kress²⁵) were targeted with the constructs indicated in Fig. 1a as described previously². Diphtheria toxin², or *PGK-DTA-bpA* (a gift from P. Soriano) and *PGK-puromycin-bpA* (a gift from A. Bradley) was used for selection in embryonic stem cells. The enhanced green fluorescent protein (EGFP) contains a Val164Ala mutation in pEGFP-C1 (Clontech; a gift from H. LeMouellic) and is followed by an SV40pA sequence. Both *Myf5^{nlacZ}* and *Myf5^{GFP-P}* comprise a *Scal-BsrYI* deletion in exon 1 (122 amino acids) and thus lack the bHLH domain, whereas the floxed *Neo-Hygro* cassette in *Myf5^{NH}* was an insertion at the *Scal* site in exon 1 (residue 13). The remaining structure of *Myf5* is intact in all alleles. The two initiator ATG codons of *Myf5* in *Myf5^{NH}* and *Myf5^{GFP-P}* were mutated (ATG to GTG). The first ATG codons suitable for translation in *Myf5^{loxP}* (residue 69, non-Kozak; residue 82, Kozak) are located in the chromatin remodelling (amino acids 39–78) and basic domains. *Myf5^{GFP-P/GFP-P}*, *Myf5^{loxP/loxP}* and *Myf5^{nlacZ/loxP}* were viable and were reduced in size. *Myod* (provided by M. Rudnicki²) and *Myf5* (129sv/C57BL6/DBA2) mutant mice were interbred to obtain double-mutant mice. A ubiquitously expressing PGKCre deleter strain (provided by Y. Lallemand²⁶) was used to generate *Myf5^{loxP}* and *Mrf4^{nlacZ-P/+}:Myf5^{loxP/+}*. *Mrf4* was targeted with an *nlacZ* reporter in embryonic stem cells previously targeted with, and *in cis* to, *Myf5^{NH}* (data not shown). All embryos and mice were genotyped by Southern blotting or polymerase chain reaction (Supplementary Methods), and double mutants were genotyped again after sectioning. *Myf5^{nlacZ}* and *Myf5^{GFP-P}* allelic combinations were also phenotyped for mislocalized MPCs by staining with 5-bromo-4-chloro-3-indolyl β -D-galactopyranoside (X-Gal)²⁷ and GFP epifluorescence, respectively.

Processing of embryos and detection of transcripts and protein

Embryos were processed for WM-ISH or immunohistochemistry¹³. Riboprobes for *Myf5*, *Mrf4* and *Myod* were as described previously^{13,28}. An *Mrf4* riboprobe spanning 421 base pairs (*KpnI-NdeI*) was also used. For comparative WM-ISH experiments, age-matched and litter-matched embryos were used with independent probe sets and litters, and the ISH reactions were stopped at the same time. For protein detection by western blotting, the head and heart of E10.75 embryos were removed and lysates were prepared²⁹. Primary polyclonal anti-Myf5 antibodies (dilution 1:500, carboxy-terminal specific; Santa Cruz) and secondary anti-rabbit peroxidase-conjugated secondary antibodies (dilution 1:40,000; Pierce) were applied and filters were treated with SuperSignal WestPico chemiluminescent substrate (Pierce).

Explant cultures and immunofluorescence

Micromass cultures were made as described previously¹³, and time-lapse studies are described in Supplementary Methods. For immunofluorescence²⁷, two antibodies were used for Myf5 (polyclonal C-terminal²⁹, and a polyclonal, dilution 1:800; Santa Cruz), myogenin (F5D, monoclonal, dilution 1:5; Developmental Studies Hybridoma Bank, NIDHD), Myod (monoclonal, dilution 1:100; Dako), myosin heavy chain (polyclonal, dilution 1:400; provided by G. Cossu) and desmin (monoclonal, dilution 1:100; Dako). Secondary antibodies were AlexaFluor 488 and 594 (Molecular Probes). For whole mount antibody staining see Supplementary Methods and <http://www.pasteur.fr/recherche/unites/Csd>. Images were taken either with a Zeiss Axiocam or Nikon numerical camera and a Zeiss fluorescent SZ11 or Axiophot microscope. Images were assembled in Adobe Photoshop and Indesign.

Received 13 March; accepted 23 July 2004; doi:10.1038/nature02876.

1. Tajbakhsh, S. Stem cells to tissue: molecular, cellular and anatomical heterogeneity in skeletal muscle. *Curr. Opin. Genet. Dev.* **13**, 412–422 (2003).
2. Rudnicki, M. A. *et al.* MyoD or myf-5 is required for the formation of skeletal muscle. *Cell* **75**, 1351–1359 (1993).
3. Kaul, A., Koster, M., Neuhaus, H. & Braun, T. Myf-5 revisited: loss of early myotome formation does not lead to a rib phenotype in homozygous myf-5 mutant mice. *Cell* **102**, 17–19 (2000).
4. Kablar, B., Krastel, K., Tajbakhsh, S. & Rudnicki, M. A. Myf5 and MyoD activation define independent myogenic compartments during embryonic development. *Dev. Biol.* **258**, 307–318 (2003).
5. Tajbakhsh, S., Rocancourt, D. & Buckingham, M. Muscle progenitor cells failing to respond to positional cues adopt non-myogenic fates in myf-5 null mice. *Nature* **384**, 266–270 (1996).
6. Kablar, B. *et al.* Myogenic determination occurs independently in somites and limb buds. *Dev. Biol.* **206**, 219–231 (1999).
7. Tajbakhsh, S. & Buckingham, M. The birth of muscle progenitor cells in the mouse: spatiotemporal considerations. *Curr. Top. Dev. Biol.* **48**, 225–268 (2000).
8. Weintraub, H. *et al.* The MyoD gene family: nodal point during specification of the muscle cell lineage. *Science* **251**, 761–766 (1991).
9. Braun, T., Rudnicki, M. A., Arnold, H. H. & Jaenisch, R. Targeted inactivation of the muscle regulatory gene Myf-5 results in abnormal rib development and perinatal death. *Cell* **71**, 369–382 (1992).
10. Tallquist, M. D., Weismann, K. E., Hellstrom, M. & Soriano, P. Early myotome specification regulates PDGFA expression and axial skeleton development. *Development* **127**, 5059–5070 (2000).
11. Kablar, B. *et al.* MyoD and Myf-5 differentially regulate the development of limb versus trunk skeletal muscle. *Development* **124**, 4729–4738 (1997).
12. Summerbell, D., Halai, C. & Rigby, P. W. Expression of the myogenic regulatory factor Mrf4 precedes or is contemporaneous with that of Myf5 in the somitic bud. *Mech. Dev.* **117**, 331–335 (2002).
13. Tajbakhsh, S., Rocancourt, D., Cossu, G. & Buckingham, M. Redefining the genetic hierarchies controlling skeletal myogenesis: Pax-3 and Myf-5 act upstream of MyoD. *Cell* **89**, 127–138 (1997).
14. Nabeshima, Y. *et al.* Myogenin gene disruption results in perinatal lethality because of severe muscle defect. *Nature* **364**, 532–535 (1993).
15. Rawls, A. *et al.* Overlapping functions of the myogenic bHLH genes MRF4 and MyoD revealed in double mutant mice. *Development* **125**, 2349–2358 (1998).
16. Moran, J. L. *et al.* Limbs move beyond the radical fringe. *Nature* **399**, 742–743 (1999).

17. Pham, C. T., MacIvor, D. M., Hug, B. A., Heusel, J. W. & Ley, T. J. Long-range disruption of gene expression by a selectable marker cassette. *Proc. Natl Acad. Sci. USA* **93**, 13090–13095 (1996).
18. Olson, E. N., Arnold, H.-H., Rigby, P. W. J. & Wold, B. J. Know your neighbors: three phenotypes in null mutants of the myogenic bHLH gene *MRF4*. *Cell* **85**, 1–4 (1996).
19. Hadchouel, J. *et al.* Modular long-range regulation of Myf5 reveals unexpected heterogeneity between skeletal muscles in the mouse embryo. *Development* **127**, 4455–4467 (2000).
20. Carvajal, J. J., Cox, D., Summerbell, D. & Rigby, P. W. A BAC transgenic analysis of the *Mrf4/Myf5* locus reveals interdigitated elements that control activation and maintenance of gene expression during muscle development. *Development* **128**, 1857–1868 (2001).
21. Maroto, M. *et al.* Ectopic Pax-3 activates MyoD and Myf-5 expression in embryonic mesoderm and neural tissue. *Cell* **89**, 139–148 (1997).
22. Lu, J. R. *et al.* Control of facial muscle development by MyoR and capsulin. *Science* **298**, 2378–2381 (2002).
23. Wang, Y. & Jaenisch, R. Myogenin can substitute for Myf5 in promoting myogenesis but less efficiently. *Development* **124**, 2507–2513 (1997).
24. Bergstrom, D. A. & Tapscott, S. J. Molecular distinction between specification and differentiation in the myogenic basic helix–loop–helix transcription factor family. *Mol. Cell. Biol.* **21**, 2404–2412 (2001).
25. Kress, C., Vandormael-Pournin, S., Baldacci, P., Cohen-Tannoudji, M. & Babinet, C. Nonpermissiveness for mouse embryonic stem (ES) cell derivation circumvented by a single backcross to 129/Sv strain: establishment of ES cell lines bearing the Omd conditional lethal mutation. *Mamm. Genome* **9**, 998–1001 (1998).
26. Lallemand, Y., Luria, V., Haffner-Krausz, R. & Lonai, P. Maternally expressed PGK-Cre transgene as a tool for early and uniform activation of the Cre site-specific recombinase. *Transgenic Res.* **7**, 105–112 (1998).
27. Tajbakhsh, S. & Buckingham, M. E. Lineage restriction of the myogenic conversion factor myf-5 in the brain. *Development* **121**, 4077–4083 (1995).
28. Bober, E. *et al.* The muscle regulatory gene, Myf-6, has a biphasic pattern of expression during early mouse development. *J. Cell Biol.* **113**, 1255–1265 (1991).
29. Daubas, P., Tajbakhsh, S., Hadchouel, J., Primig, M. & Buckingham, M. Myf5 is a novel early axonal marker in the mouse brain and is subjected to post-transcriptional regulation in neurons. *Development* **127**, 319–331 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. Shorte, E. Perret and P. Roux at the Pasteur Dynamic Imaging Center, and C. Cimper for assistance. The laboratories of S.T. and M.B. were funded by the Pasteur Institute, CNRS, AFM, ACI Integrative Biology programme of the French Ministry. M.B. was funded by the European Union. S.T. was also funded by grants from the HFSP, Association pour la Recherche sur le Cancer. B.G.-M. received a fellowship from the ARC, and V.S. a grant from the Fondation pour la Recherche Médicale and Pasteur Institute. L.K.-D. was funded by the HFSP, the EU and Pasteur Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.T. (shaht@pasteur.fr).

Exogenous control of mammalian gene expression through modulation of RNA self-cleavage

Laising Yen¹, Jennifer Svendsen¹, Jeng-Shin Lee¹, John T. Gray¹, Maxime Magnier¹, Takashi Baba², Robert J. D'Amato² & Richard C. Mulligan¹

¹Department of Genetics, Harvard Institute of Human Genetics, Harvard Medical School, and Division of Molecular Medicine and

²Department of Surgery, Children's Hospital, Boston, Massachusetts 02115, USA

Recent studies on the control of specific metabolic pathways in bacteria have documented the existence of entirely RNA-based mechanisms for controlling gene expression. These mechanisms involve the modulation of translation, transcription termination or RNA self-cleavage through the direct interaction of specific intracellular metabolites and RNA sequences^{1–4}. Here we show that an analogous RNA-based gene regulation system can effectively be designed for mammalian cells via the incorporation of sequences encoding self-cleaving RNA motifs⁵ into the transcriptional unit of a gene or vector. When correctly positioned, the sequences lead to potent inhibition of gene or vector expression,

owing to the spontaneous cleavage of the RNA transcript. Administration of either oligonucleotides complementary to regions of the self-cleaving motif or a specific small molecule results in the efficient induction of gene expression, owing to inhibition of self-cleavage of the messenger RNA. Efficient regulation of transgene expression is shown in a variety of mammalian cell lines and live animals. In conjunction with other emerging technologies⁶, this methodology may be particularly applicable to the development of gene regulation systems tailored to any small inducer molecule, and provide a novel means of biological sensing *in vivo* that may have an important application in the regulated delivery of protein therapeutics.

The general strategy for controlling gene expression through modulation of RNA processing is shown in Fig. 1a. The approach is critically dependent on both the ability of a specific self-cleaving ribozyme to efficiently cleave (>99%) an mRNA molecule in which it is embedded, and the availability of a small molecule capable of efficiently inhibiting self-cleavage of the cis-acting ribozyme within an intracellular milieu. Candidate ribozyme sequences capable of efficient cleavage in mammalian cells were identified using a transient transfection assay which makes use of a standard mammalian expression vector⁷ which encodes a LacZ reporter (Fig. 1b). Candidate ribozyme sequences were introduced into one of a number of different locations within the vector transcriptional unit, and, in parallel, the corresponding inactive mutant ribozymes were introduced into the same sites to measure the efficiency of reduction of reporter gene expression. After transfection of human embryonic kidney (HEK) 293 cells with these vectors, protein extracts were prepared from cells and the β -gal level quantified. Cleavage activity of the functional ribozyme in cells was measured as 'fold' suppression in reporter gene expression relative to the vector with inactive ribozyme.

A large number of different ribozyme encoding motifs were chosen for analysis, including un-manipulated natural ribozyme sequences, ribozymes shown to function in mammalian cells and ribozymes engineered by others to possess specific biochemical or catalytic properties *in vitro* (for example, high K_{cat} , low Mg^{2+} requirement, and so on; see Supplementary Table 1). Although the vast majority of ribozymes tested did not appreciably affect reporter expression, as reflected by near equal expression of LacZ by vectors encoding functional and inactive ribozymes (defined as fold = 1), two ribozyme motifs were identified which did seem to function effectively: the hammerhead Pst-3 ribozyme derived from the *Dolichopoda* cave cricket⁸ (Fig. 1c; 13-fold difference between functional and inactive ribozyme) and the hammerhead Sm1 ribozyme derived from the trematode *Schistosoma mansoni*⁹ (Fig. 1d) in which the tetraloop 5'UUCG 3' was grafted onto an extended stem III (19-fold difference). Interestingly, both ribozymes possess unique structures relative to the other hammerhead ribozymes tested, in that they contain an extended stem-I with an internal loop (see Fig. 1c, d).

Based on its apparent higher level of self-cleavage activity, the Sm1 ribozyme was chosen for further study and manipulation. In an effort to improve the efficiency of the Sm1 ribozyme self-cleavage activity, a series of modifications of the Sm1 ribozyme structure were made and evaluated. As shown in Fig. 1d, specific modification at nucleotide 7 (C to U) in the conserved catalytic core¹⁰, and changes in distal stem III led to a significant increase in the extent of self-cleavage (the modified Sm1 ribozyme was termed N73, up to 62-fold difference between functional and inactive ribozyme). Transfer of the N73 ribozyme from position A to E of the vector enhanced the activity to 225-fold (this ribozyme was termed N79 and was used in the subsequent studies). Additional modifications of stem I near the catalytic core and near the restriction insertion site led to further increases in activity, resulting ultimately in an overall 1,401-fold difference in expression levels between functional versus inactive ribozyme (Fig. 1e).

In addition to modifications that led to improved self-cleavage activity, several modifications, notably those involving the shortening of stem I (Fig. 1f) and alteration of nucleotides within the internal loop of stem I (Fig. 1g), dramatically reduced the level of self-cleavage. Interestingly, neither of those two modifications affect conserved core sequences known to be required for ribozyme cleavage *in vitro*¹¹. Under standard *in vitro* conditions of 10 mM Mg^{2+} , measurement of the catalytic activity of N79 ribozyme and N107(U to G) ribozyme (which carries a single base U to G change in the loop I and is inactive in mammalian cells, Fig. 1g) indicated that both enzymes were equally functional *in vitro* (K_{obs} values of 0.84 min⁻¹ and 1.06 min⁻¹, respectively; see Supplementary Methods). Intriguingly, determination of the cleavage rate under low Mg^{2+} conditions (0.5 mM Mg^{2+}) indicated that only the N79 enzyme possessed significant activity (N79 K_{obs} = 0.84 min⁻¹ versus N107(U to G) K_{obs} = 0.014 min⁻¹). These results suggest that sequences within the unique loop structure of stem-I of the Sm1 ribozyme may enable efficient self-cleavage in mammalian cells in part because they facilitate self-cleavage at physiological Mg^{2+} concentrations. Consistent with this idea, measurement of the K_{obs} of another well-characterized hammerhead ribozyme¹² ('McSwiggen's hammerhead'; see Supplementary Table 1) previously shown by others to function in mammalian cells, but shown in our transfection assay to possess no appreciable activity, indicated that significant *in vitro* cleavage activity occurred only under high Mg^{2+} conditions (0.32 min⁻¹ at 10 mM Mg^{2+} versus 0.015 min⁻¹ at 0.5 mM Mg^{2+}). However, to what extent the ability to function at low Mg^{2+} concentration *per se* contributes to the ability of a ribozyme to function efficiently in mammalian cells remains unclear, because several ribozymes engineered by others to efficiently function *in vitro* under low Mg^{2+} conditions^{13,14} were among the many ribozymes that were found to be non-functional in our transient transfection assay (Supplementary Table 1).

In addition to functioning in HEK293 cells within the context of the original cytomegalovirus (CMV)-based pMD vector tested, the N79 ribozyme also dramatically reduced β -gal expression in a variety of other commonly used cell lines after transfection (Fig. 2a). In addition, the N79 ribozyme was able to function efficiently when placed within other transcriptional units which made use of different promoters and a different reporter gene (enhanced green fluorescent protein, eGFP) (Fig. 2b). As we had observed in the primary screen of different ribozyme motifs for activity in mammalian cells, N79 was able to function when placed in several, but not all locations of the pMD transcriptional unit, albeit at different efficiencies (Fig. 2c). Importantly, placement of two ribozyme sequences in tandem, in some cases (for example, position E) led to a dramatic suppression of reporter gene expression (Fig. 2c).

An essential requirement for the development of a gene regulation system based on modulation of self-cleavage activity is the availability of small inducer molecules capable of efficient inhibition of ribozyme activity in mammalian cells. In an effort to identify such molecules, we first surveyed a large number of common antibiotics that had been shown to inhibit ribozyme cleavage *in vitro*¹⁵⁻²⁰. In no case was significant inhibition of self-cleavage by these antibiotics observed in our transient transfection assay (data not shown). We next surveyed the ability of different types of antisense oligonucleotides²¹ to inhibit ribozyme cleavage (see Fig. 1e for targeted sequence). Whereas transfection with peptide nucleic acids, locked nucleic acids and 'grip' nucleic acids had no measurable effects on self-cleavage activity, transfection with phosphorothiolate, 2'-O-methyl and phosphorothiolate 2'-O-methyl derived RNAs led to modest inhibition of self-cleavage (<tenfold induction). In contrast, transfection of a morpholino oligonucleotide²² led to a strong inhibition of ribozyme self-cleavage, as demonstrated by a dramatic increase in reporter gene expression (Fig. 3a). The 'fold' induction afforded by this method was somewhat variable

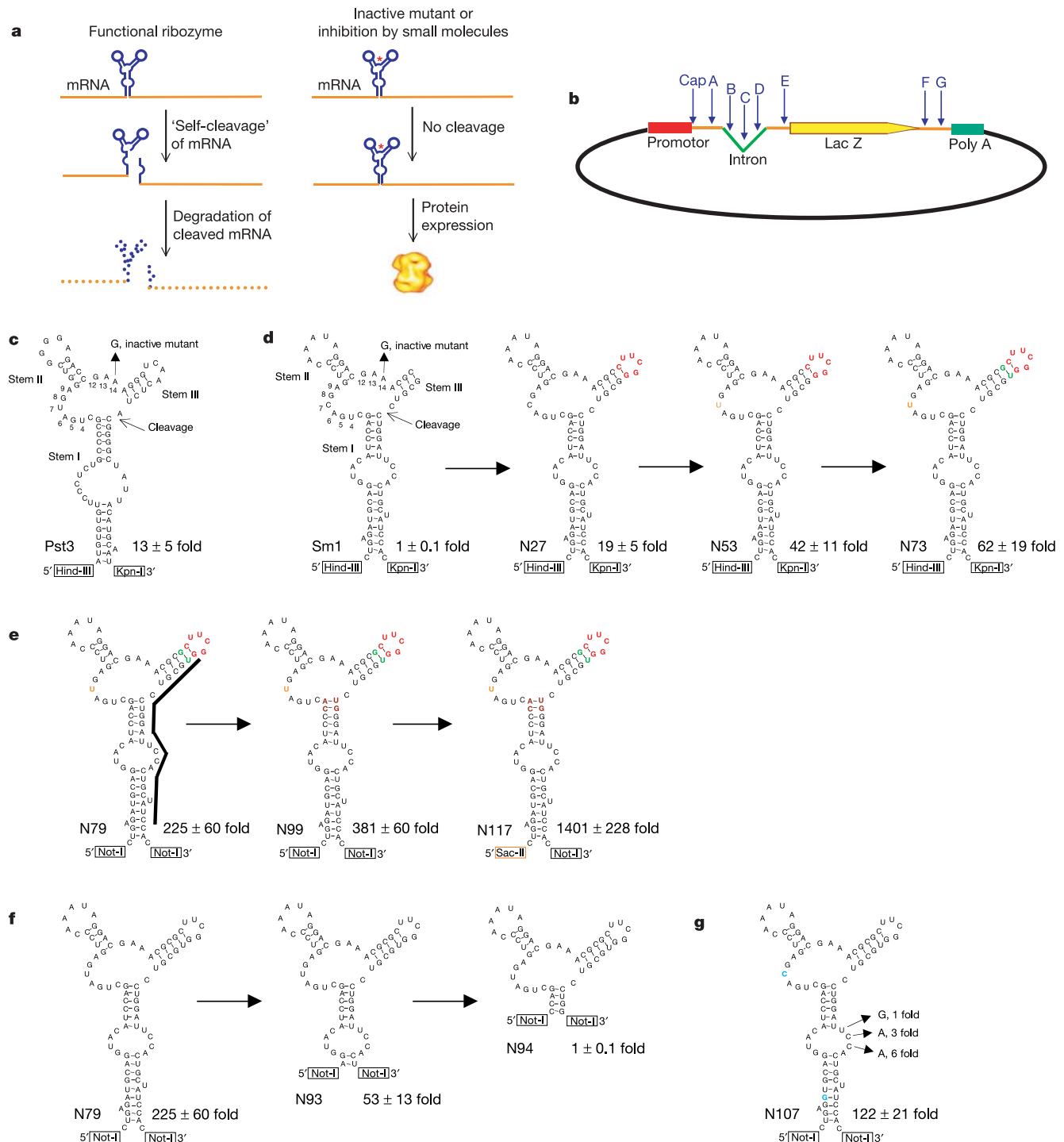


Figure 1 Strategy for controlling gene expression through the modulation of RNA self-cleavage and optimization of Schistosoma Sm1 ribozyme self-cleavage activity. **a**, When a cis-acting hammerhead ribozyme is embedded in the mRNA, self-cleavage leads to the destruction of the mRNA and the absence of gene expression. However, an inactive mutant or the administration of specific inhibitors of the ribozyme leads to the generation of intact mRNA's and protein expression. **b**, The reporter gene expression vector pMD used for transfection assay, and the positions where ribozyme were placed. Cap site at the very beginning of the mRNA; A site upstream of the intron; B, C, and D site in the intron; E site immediately upstream of the translation start; F and G sites in the 3'-untranslated region. **c-g**, Optimization of Schistosoma Sm1 activity. Ribozyme inserted at position A of pMD vector (**c, d**) and at position E of vector (**e-g**). Corresponding inactive mutants

contained an A14 to G substitution. Name of ribozyme is shown at left; cleavage activity at right. Cricket Pst3 motif (**c**). Nucleotide numbering follows nomenclature of Hertel *et al.*¹⁰. Schistosoma Sm1 motif (**d**). The original Sm1 lacked loop III and exhibited no activity in cells. Nucleotides depicted in colour reflect changes made to Sm1. Changes in stem I of N79 near the core or near the restriction insertion site further enhanced activity (**e**). Black line identifies the sequence targeted by antisense morpholino oligonucleotides. Shortening of stem I vastly reduced ribozyme cleavage activity (**f**). Single nucleotide changes in loop I of the N107 ribozyme decreased its activity dramatically (**g**). The N107 ribozyme is a variant of N79 in which two 'AUG' were replaced by GUG and ACG to eliminate the potential start codons. The mean $\pm$ s.d. of at least four independent measurements is shown.

from experiment to experiment (110–2000-fold in the case of the double N79 construct), probably because of the variability of efficiency of delivery and toxicity associated with transfection of the oligonucleotide. However, the extent of induction of gene expression was nonetheless comparable to that achieved with other gene regulation systems, and clearly represents a range that would be useful for a variety of experimental and clinical settings. In some cases, the absolute levels of gene expression achieved after morpholino administration approached 50% of the theoretical maximum induction possible (that is, the level of gene expression produced by the inactive ribozyme), suggesting that ribozyme cleavage can be very efficiently inhibited in mammalian cells.

To identify small molecules capable of inhibiting ribozyme self-cleavage, we generated stable cell lines carrying an integrated expression construct in which a luciferase reporter was placed under the control of two copies of the N79 ribozyme, and made use of the cells in high-throughput screening studies (L.Y., M.M., Y. Tang, R. Weissleder, B. R. Stockwell and R.C.M., unpublished observations). Of the several compounds identified, toyocamycin²³ (a nucleoside analogue) was found to be the most potent inhibitor of ribozyme function. As shown in Fig. 3b, administration of 1.5 μ M toyocamycin to the cells led to a dramatic increase in luciferase protein expression. A parallel analysis of luciferase mRNA expression demonstrated that in the absence of drug, little if any luciferase mRNA was produced in the nucleus or cytoplasm, whereas in drug-treated cells, the amount of luciferase mRNA was increased to a level comparable to that of cells carrying inactive ribozyme (Fig. 3c). The lack of detectable mRNA in untreated cells suggests that cleaved mRNAs are rapidly degraded, presumably

because the cleaved fragments lack conventional sequences at the ends of mRNAs. In contrast to toyocamycin, a related compound, adenosine, possessed no ability to inhibit ribozyme self-cleavage (data not shown). Although the above assays indicate that self-cleavage is highly efficient in the stable cell line, sensitive measurement of luciferase activity by photon emission indicates that there is, nevertheless, an extremely low but detectable amount of luciferase expression above the background emission level of control cells that carry no luciferase gene (see legend to Fig. 3).

Having documented the ability of the ribozyme-based regulation

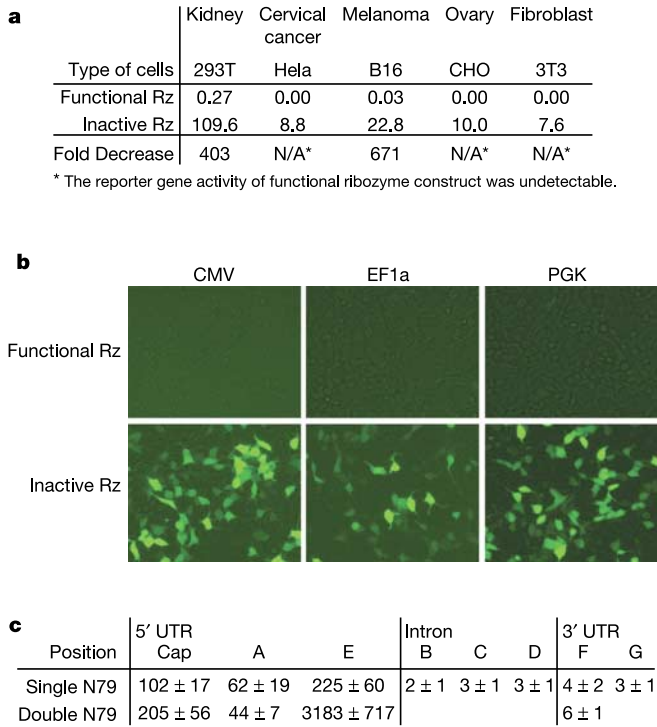


Figure 2 Efficient self-cleavage can occur in different cells, with different vectors and with ribozyme sequences positioned in different locations. **a**, N79 functioned efficiently in a variety of cell types. Numbers are the measurements of β -gal expressed. **b**, N79 functioned efficiently with CMV, EF1a and PGK promoters in a vector encoding the eGFP reporter. **c**, N79 functioned efficiently at some but not all positions within the vector. Numbers are 'fold decrease' in β -gal expression (functional verses inactive ribozyme). The mean $\pm$ s.d. from at least four independent measurements is shown.

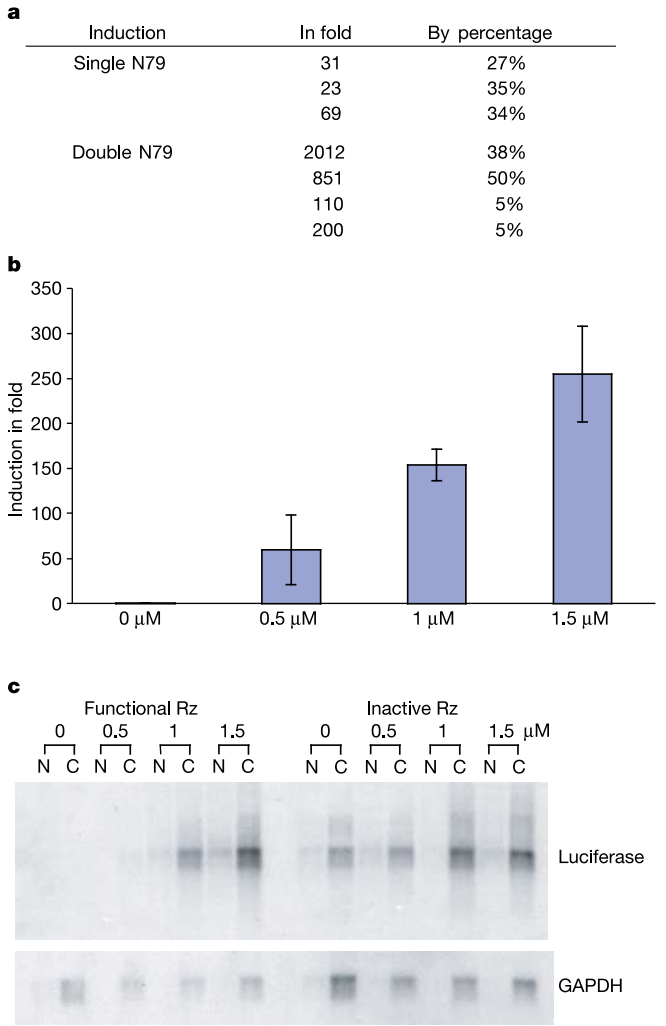


Figure 3 Induction of gene expression in cultured cells through inhibition of ribozyme self-cleavage. **a**, Effect of morpholino oligonucleotide on N79 ribozyme in transient transfection assay. Induction was measured by 'fold increase' in β -gal expression with or without morpholino application. The level of induction is also shown as a percentage relative to the expression level of inactive ribozyme. The target for the oligonucleotide is shown in Fig. 1e. **b**, Induction of luciferase expression by toyocamycin in a stable cell line carrying an expression construct containing a double N79 ribozyme. Cells were treated with toyocamycin for 24 h at dosage of 0, 0.5, 1 and 1.5 μ M (toxic effects were observed at concentrations higher than 1.5 μ M). Quantitative measurements of luciferase activity revealed emission of 1555, 66774, 242546 and 377655 photons per second per 1,000 cells respectively, as compared to a background emission of 121 from cells carrying no luciferase gene. Values are mean $\pm$ s.d. from at least four independent measurements. **c**, Induction by toyocamycin at the RNA level as shown by northern blot analyses. Experimental conditions were similar to that of **b**. RNA was purified from the nucleus (N) or the cytoplasm (C) after 24 h of treatment.

system to function in mammalian cell culture, an important issue was whether the intracellular machinery and 'environment' necessary for self-cleavage was operable in primary cells *in vivo*. To address this issue, we generated recombinant adeno-associated virus (AAV) genomes carrying transcription units derived from pMD ribozyme-luciferase vectors (possessing two copies of functional or inactive N79 ribozyme at the E position), and prepared high titre virus possessing the host range of AAV serotype 5. The two viruses were then used to inject nude mice sub-retinally as described in Supplementary Methods. To provide a means of normalization for differences in the capacity to measure luciferase activity on different days (such as, variations due to the delivery of the luciferin substrate), all animals were also injected in the hamstring muscles of the hind limb with AAV viruses carrying the inactive N79 ribozymes. After 21 days, the injected animals were imaged for luciferase gene expression using the Xenogen IVIS imager (which provides a quantitative measure of luciferase expression based on single-photon detection)²⁴. Immediately after imaging, cohorts of mice injected with virus carrying the functional ribozyme sequences were implanted under the dorsal skin with seven day 'time-release' pellets of either toyocamycin or adenosine, whereas another cohort of mice injected with virus carrying the inactive ribozyme sequences were implanted with toyocamycin pellets. Two days later, all animals were then imaged for luciferase expression.

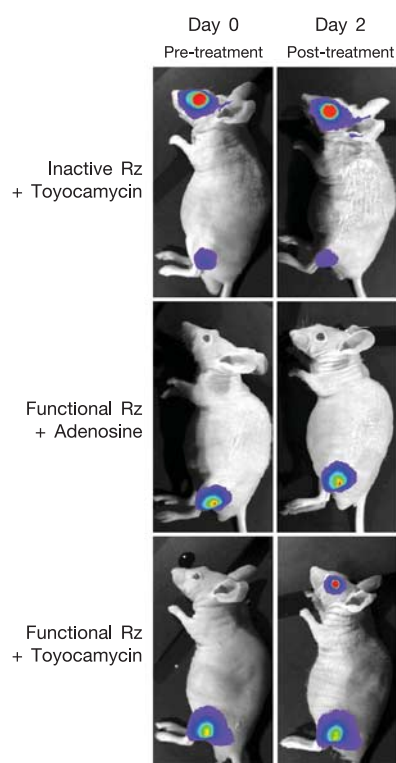


Figure 4 Effective control of gene expression *in vivo* using ribozyme-based gene regulation system. Upper panels show animal in which retina was injected with AAV carrying double inactive N79 and treated with toyocamycin. Middle panels show animal in which retina was injected with AAV carrying double functional N79 and treated with adenosine. Lower panels show same as middle panel but treated with toyocamycin. A strong induction in luciferase expression at day 2 was observed in lower panel. Animals were also injected in the hamstring muscles of the hind limb with AAV carrying double inactive N79 as internal control. AAV injection was done 3 weeks before the first imaging day (day 0). Drug-releasing pellets were implanted subcutaneously in the dorsal neck immediately after the first imaging, and released the drug over 7 days. The toyocamycin pellet contains 10 µg of drug; adenosine pellet 50 µg.

Representative images of mice in each treatment group, taken before and after drug treatment, are shown in Fig. 4. The images demonstrate that, as expected, mice injected with virus carrying inactive ribozymes showed robust luciferase expression in the retina, and the expression was independent of the administration of toyocamycin (Fig. 4, upper panel). Mice injected with virus carrying two functional ribozymes and implanted with adenosine pellets showed little if any gene expression before or after adenosine treatment (Fig. 4, middle panel), consistent with the inability of adenosine to inhibit ribozyme self-cleavage. Importantly, mice injected with virus carrying the functional ribozymes showed readily detectable expression only after toyocamycin treatment (Fig. 4, lower panel). Quantification of the photon output indicated gene expression was 'induced' 39, 185 and 191-fold in the three mice infected with virus carrying the functional ribozymes and treated with toyocamycin. In the last case (animal showing 191-fold induction) the induced gene expression reached a level within 40% of the gene expression of virus carrying inactive ribozymes. These results indicate that significant ribozyme-mediated gene regulation can be accomplished in the *in vivo* setting. Whereas the retina may be particularly accessible to 'inducer' due to its extensive vascularization, we have shown in preliminary experiments that gene regulation can be accomplished at a number of other anatomical sites *in vivo* (unpublished observations).

Overall, the studies reported here provide an important 'proof-of-principle' for gene regulation strategies based on the modulation of RNA processing. Specifically, the fact that efficient ribozyme self-cleavage can occur in a variety of mammalian cell lines and in primary cells *in vivo*, suggests that mammalian cells may in general be 'permissive' for efficient ribozyme self-cleavage. Therefore, ribozyme-based regulation systems may be generally applicable to the manipulation of gene expression in cells and animals. In addition to the implications for the development of gene regulation strategies, the studies also provide a compelling rationale for determining whether 'naturally occurring' RNA-only mechanisms for gene regulation exist in mammalian cells.

The most commonly used systems for controlling gene expression, that rely on the regulation of transcription^{25–28}, have proved to be extremely powerful experimental tools. However, despite their utility, such systems possess at least some practical and theoretical limitations because of their reliance on chimaeric transcriptional transactivators and specialized promoter elements. These limitations include: (1) the need to co-introduce expression constructs for both the transactivator and the transgene to be regulated; (2) the potential toxicities due to expression of a chimaeric transactivator; (3) difficulties in application of such systems to the regulation of endogenous cellular genes due to the requirement of a specialized promoter; and (4) the limited number of small inducer molecules available for experimental and therapeutic applications.

In contrast to systems based on the regulation of transcription, the ribozyme-based system we have described does not require the expression of any protein transactivator products and is not dependent on the use of any specialized promoter elements. Therefore, in theory, this system represents a 'portable' regulation system that could be 'embedded' into any endogenous gene or engineered vector transcription unit. Although the two inhibitors of ribozyme self-cleavage we have described may not be ideal for many experimental applications, it is likely that additional inducers with more desirable pharmacokinetic properties and toxicity profiles can be identified through high-throughput screening efforts and further evaluation of specific antisense oligonucleotides and *in vivo* delivery methods to cells. In addition, recent studies have shown that it is possible to generate ribozymes whose *in vitro* self-cleavage activity is controlled by a specific ligand, either by the 'judicious' linkage of RNA aptamer sequences to specific regions of hammerhead ribozyme²⁹, or through the use of *in vitro* evolution technologies³⁰.

Application of these technologies to the strategy for controlling gene expression described here should make it possible in the future to tailor specific ribozyme-based gene regulation systems to any small molecule ligand. Such an approach would provide a general methodology for developing gene regulation systems which rely on ligands with desirable and/or specific pharmacokinetic properties. In addition, the combined technologies should provide the means to independently and simultaneously control the expression of multiple gene products, and to express gene products in response to the concentration of any intracellular molecule or combinations of molecules. Such a form of biological sensing could have broad experimental and therapeutic applications. □

Methods

Transient reporter expression assay

Plasmids carrying different ribozymes were transfected into HEK293T cells, which were lysed after 24 h. The cell extracts were incubated with ONPG (o-nitrophenyl- β -D-galactopyranoside), and the amount of β -galactosidase was measured by quantifying the processed ONPG using a luminometer.

Catalytic rate measurement

Ribozymes were generated by *in vitro* transcription in the presence of 50 μ M blocking antisense oligonucleotides. Full length ribozymes were purified and the cleavage rate determined in 50 mM Tris-HCl, pH 7.5 at 23 °C. K_{obs} was calculated according to the equation $F_t = F_0 + F_{\infty}(1 - e^{-kt})$, where F_0 and F_{∞} are the cleavage fractions at zero time and at the reaction endpoint, and k is the first order rate constant of cleavage (K_{obs}).

Non-invasive bioluminescent imaging

Before imaging, the anaesthetized mice were injected with 150 μ l of luciferin (30 mg ml⁻¹) and the pupils were moistened and dilated with 1% tropicamide. A series of bioluminescent images were taken for up to 30 min using the Xenogen IVIS imager. Photon output was quantified at the plateau of the time course using the LivingImage software. Induction in fold was calculated based on the photon output in the retina before and after drug treatment, and was normalized to the photon output from the leg muscles. See Supplementary Information for more detailed Methods.

Received 25 May; accepted 14 July 2004; doi:10.1038/nature02844.

- Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952–956 (2002).
- Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. & Breaker, R. R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**, 281–286 (2004).
- Mandal, M. & Breaker, R. R. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature Struct. Mol. Biol.* **11**, 29–35 (2004).
- Cech, T. R. RNA finds a simpler way. *Nature* **428**, 263–264 (2004).
- Cech, T. R. Nobel lecture. Self-splicing and enzymatic activity of an intervening sequence RNA from *Tetrahymena*. *Biosci. Rep.* **10**, 239–261 (1990).
- Silverman, S. K. Rube Goldberg goes (ribo)nuclear? Molecular switches and sensors made from RNA. *RNA* **9**, 377–383 (2003).
- Ory, D. S., Neugeboren, B. A. & Mulligan, R. C. A stable human-derived packaging cell line for production of high titer retrovirus/vesicular stomatitis virus G pseudotypes. *Proc. Natl Acad. Sci. USA* **93**, 11400–11406 (1996).
- Rojas, A. A. *et al.* Hammerhead-mediated processing of satellite pD500 family transcripts from *Dolichopoda* cave crickets. *Nucleic Acids Res.* **28**, 4037–4043 (2000).
- Ferbeyre, G., Smith, J. M. & Cedergren, R. *Schistosoma* satellite DNA encodes active hammerhead ribozymes. *Mol. Cell. Biol.* **18**, 3880–3888 (1998).
- Hertel, K. J. *et al.* Numbering system for the hammerhead. *Nucleic Acids Res.* **20**, 3252 (1992).
- Ruffner, D. E., Stormo, G. D. & Uhlenbeck, O. C. Sequence requirements of the hammerhead RNA self-cleavage reaction. *Biochemistry* **29**, 10695–10702 (1990).
- Chowrira, B. M., Pavco, P. A. & McSwiggen, J. A. *In vitro* and *in vivo* comparison of hammerhead, hairpin and hepatitis delta virus self-processing ribozyme cassettes. *J. Biol. Chem.* **269**, 25856–25864 (1994).
- Zillmann, M., Limauro, S. E. & Goodchild, J. *In vitro* optimization of truncated stem-loop II variants of the hammerhead ribozyme for cleavage in low concentrations of magnesium under non-turnover conditions. *RNA* **3**, 734–747 (1997).
- Conaty, J., Hendry, P. & Lockett, T. Selected classes of minimised hammerhead ribozyme have very high cleavage rates at low Mg²⁺ concentration. *Nucleic Acids Res.* **27**, 2400–2407 (1999).
- Hermann, T. & Westhof, E. Aminoglycoside binding to the hammerhead ribozyme: a general model for the interaction of cationic antibiotics with RNA. *J. Mol. Biol.* **276**, 903–912 (1998).
- Jenne, A. *et al.* Rapid identification and characterization of hammerhead-ribozyme inhibitors using fluorescence-based technology. *Nature Biotechnol.* **19**, 56–61 (2001).
- Murray, J. B. & Arnold, J. R. Antibiotic interactions with the hammerhead ribozyme: tetracyclines as a new class of hammerhead inhibitor. *Biochem. J.* **317**, 855–860 (1996).
- Stage, T. K., Hertel, K. J. & Uhlenbeck, O. C. Inhibition of the hammerhead ribozyme by neomycin. *RNA* **1**, 95–101 (1995).
- Tor, Y., Hermann, T. & Westhof, E. Deciphering RNA recognition: aminoglycoside binding to the hammerhead ribozyme. *Chem. Biol.* **5**, R277–R283 (1998).
- von Ahlen, U., Davies, J. & Schroeder, R. Antibiotic inhibition of group I ribozyme function. *Nature* **353**, 368–370 (1991).
- Braasch, D. A. & Corey, D. R. Novel antisense and peptide nucleic acid strategies for controlling gene expression. *Biochemistry* **41**, 4503–4510 (2002).

- Morcos, P. A. Achieving efficient delivery of morpholino oligos in cultured cells. *Genesis* **30**, 94–102 (2001).
- Aszalos, A., Lemanski, P., Robison, R., Davis, S. & Berk, B. Identification of antibiotic 1037 as toyocamycin. *J. Antibiot. (Tokyo)* **19**, 285 (1966).
- Contag, P. R., Olomu, I. N., Stevenson, D. K. & Contag, C. H. Bioluminescent indicators in living mammals. *Nature Med.* **4**, 245–247 (1998).
- Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl Acad. Sci. USA* **89**, 5547–5551 (1992).
- Rivera, V. M. *et al.* A humanized system for pharmacologic control of gene expression. *Nature Med.* **2**, 1028–1032 (1996).
- Suhr, S. T., Gil, E. B., Senut, M. C. & Gage, F. H. High level transactivation by a modified *Bombix* ecdysone receptor in mammalian cells without exogenous retinoid X receptor. *Proc. Natl Acad. Sci. USA* **95**, 7999–8004 (1998).
- Wang, Y., O'Malley, B. W. Jr, Tsai, S. Y. & O'Malley, B. W. A regulatory system for use in gene transfer. *Proc. Natl Acad. Sci. USA* **91**, 8180–8184 (1994).
- Breaker, R. R. Engineered allosteric ribozymes as biosensor components. *Curr. Opin. Biotechnol.* **13**, 31–39 (2002).
- Wilson, D. S. & Szostak, J. W. *In vitro* selection of functional nucleic acids. *Annu. Rev. Biochem.* **68**, 611–647 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank K. Salehi-Ashtiani and J. Szostak for helpful discussions, Y. Tang and R. Weissleder for help with imaging experiments performed during the early course of the work, and M. Chung for her technical assistance. This work was supported by grants from AMGEN and L'Association Francaise contre les Myopathies (AFM). R.C.M. is an AMGEN consultant and equity holder.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to R.M. (mulligan@receptor.med.harvard.edu).

Tropism switching in *Bordetella* bacteriophage defines a family of diversity-generating retroelements

Sergei Doulatov^{1*}, Asher Hodes^{1*}, Lixin Dai³, Neeraj Mandhana¹, Mingsun Liu¹, Rajendar Deora¹, Robert W. Simons^{1,2}, Steven Zimmerly³ & Jeff F. Miller^{1,2}

¹Department of Microbiology, Immunology and Molecular Genetics, David Geffen School of Medicine at UCLA, and

²The Molecular Biology Institute, University of California, Los Angeles, California 90095, USA

³Department of Biological Sciences, University of Calgary, Calgary, Alberta T2N 1N4, Canada

* These authors contributed equally to this work

Bordetella bacteriophages generate diversity in a gene that specifies host tropism¹. This microevolutionary adaptation is produced by a genetic element that combines the basic retroelement life cycle of transcription, reverse transcription and integration with site-directed, adenine-specific mutagenesis. Central to this process is a reverse transcriptase-mediated exchange between two repeats; one serving as a donor template (TR) and the other as a recipient of variable sequence information (VR)¹. Here we describe the genetic basis for diversity generation. The directionality of information transfer is determined by a 21-base-pair sequence present at the 3' end of VR. On the basis of patterns of marker transfer in response to variant selective pressures, we propose that a TR reverse transcriptase is mutagenized, integrated into VR as a single non-coding strand, and then partially converted to the parental VR sequence. This allows the diversity-generating system to minimize variability to the subset of bases under selection. Using the *Bordetella* phage

cassette as a signature, we have identified numerous related elements in diverse bacteria. These elements constitute a new family of retroelements with the potential to confer selective advantages to their host genomes.

Bordetella species cause respiratory infections in mammals. Their cell surface is highly variable as a result of a complex programme of gene expression mediated by the BvgAS phosphorelay, which regulates the infectious cycle^{2–5}. *Bordetella* bacteriophages encode a unique diversity-generating system that allows the use of different receptor molecules for attachment and infection^{1,6}. Phage BPP-1 preferentially infects virulent, Bvg⁺ phase *Bordetella* owing to differential expression of the outer membrane protein pertactin (Prn), which functions as the phage receptor (Fig. 1a)^{1,7–9}. At

characteristic frequencies, BPP phage give rise to tropic variants that recognize distinct surface receptors and preferentially infect avirulent, Bvg[–] phase bacteria (BMP phage) or that are indiscriminate to the Bvg phase (BIP phage). These viral parasites have thus evolved to keep pace with the dynamic surface structure displayed by their host as it travels through its infectious cycle.

Tropism switching is mediated by a variability generating cassette encoded in the phage genome (Fig. 1b). This cassette functions to introduce nucleotide substitutions at 23 sites in a 134-base-pair (bp) VR present at the 3' end of the *mtd* locus. Sites of variability in VR correspond to adenine residues in a nearby homologous TR, which is invariant and essential for tropism switching. As silent substitutions in TR are transmitted to VR during switching¹, TR supplies the raw sequence information for variability. The *brt* locus, which encodes a reverse transcriptase (RT), and *atd*, which has an unknown function, are also essential for the generation of diversity (ref. 6 and Supplementary Fig. 1). These observations suggest a mechanism in which a TR-containing transcript serves as a template for Brt-catalysed reverse transcription and adenine-specific mutagenesis, followed by replacement of VR by the TR-derived complementary DNA product in an event termed 'mutagenic homing'. Mutagenesis occurs exclusively at adenines, with no other sequence requirements. Substitution of an adenine in TR with another nucleotide abolishes variation at that position, whereas introduction of an ectopic adenine produces a novel site of heterogeneity (Supplementary Fig. 2). Tropic variants may contain anywhere from 1 to 23 substitutions and the system is theoretically capable of generating over 10¹² polypeptide sequences. Mtd, a putative tail protein, is necessary for phage morphogenesis and infectivity, and the sequence of VR within Mtd determines tropism specificity¹. Binding of a BPP-1-derived glutathione S-transferase (GST)–Mtd fusion protein to the *Bordetella* cell surface is dependent on expression of Prn, correlating with the infective properties of the parental phage (Supplementary Fig. 3). The cassette shown in Fig. 1b therefore functions to generate plasticity in a ligand–receptor interaction.

A hallmark of the diversity-generating system is the directional transfer of sequence information that accompanies mutagenesis. Although VR is highly variable, TR is maintained as an uncorrupted source of sequence information. Located at the 3' end of each repeat is a 15-bp segment consisting of G and C residues, followed by a 21-bp sequence that differs at five positions between TR and VR (Fig. 1c). We tested the possibility that these polymorphisms form part of a *cis*-acting site that determines the directionality of homing. Substitution of the 21-bp VR sequence with the corresponding TR sequence eliminated tropism switching (BPP-3'VR), whereas the reverse substitution did not (BPP-3'TR). Remarkably, using polymerase chain reaction (PCR)-based *in vitro* variability assays, we observed that the latter substitution resulted in the generation of variability in TR as well as VR (Fig. 1c), an event that is never observed in wild-type phage. Variability occurred solely at positions occupied by adenine residues in the parental TR, indicating that the basic mechanism of mutagenesis was retained. Furthermore, the pattern of mutations observed in different BPP-3'VR phage indicated that TR was the source of both TR and VR variability (Fig. 1d). These observations demonstrate that the sequence designated IMH (initiation of mutagenic homing; Fig. 1c) is responsible for the directional transfer of sequence information. In contrast, deletion analysis indicated that the 5' boundary of information transfer is established by the extent of homology between VR and TR (Supplementary Fig. 4). Thus, it seems that the 'genetic logic' of the wild-type system is to prevent corruption of TR while limiting variability to VR.

To generate a more comprehensive view of the mutagenic homing reaction, we designed a genetic strategy for tracking events that give rise to sequence variants based on the observation that conservative nucleotide substitutions in TR are incorporated into VRs of phages

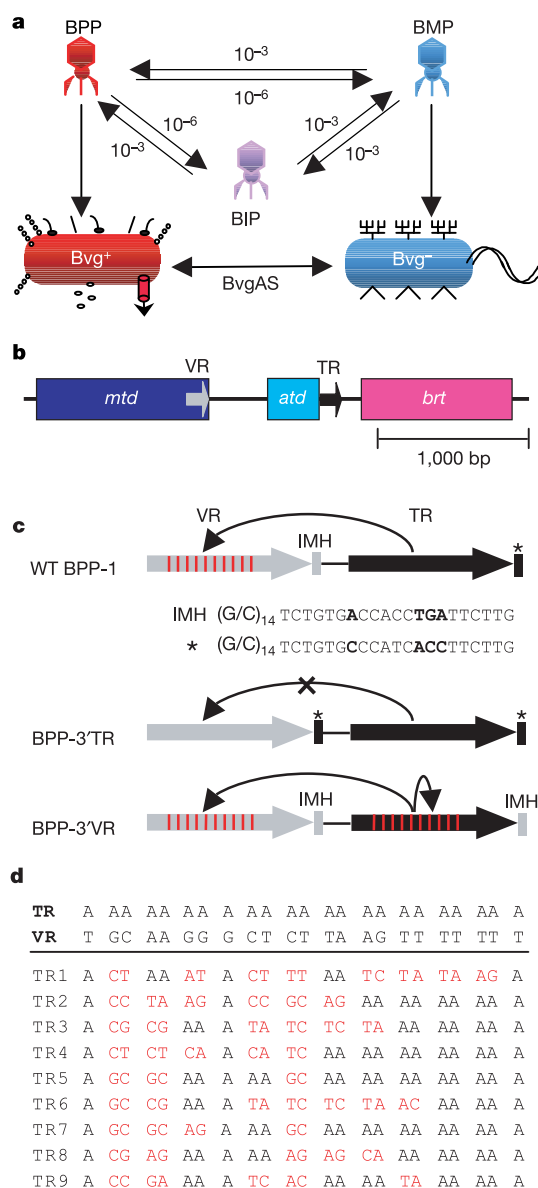


Figure 1 Tropism switching by *Bordetella* bacteriophage. **a**, Specificities and switching frequencies are depicted above the *B. bronchiseptica* BvgAS-mediated phase transition⁴. **b**, Components of the variability generating cassette. **c**, Directional transfer of sequence information is directed by the IMH sequence. Red lines indicate adenine-specific mutagenesis. WT, wild type. **d**, TR adenines are shown at the top followed by the corresponding nucleotides in the parental VR. TR1–9 are TR sequences derived from *in vitro* variability assays performed on phage BPP-3'VR. Red nucleotides show positions that varied.

that have switched tropism. By introducing multiple synonymous substitutions positioned along TR, the portion of TR transferred during a switching event can be determined by recording the pattern of substitutions appearing in VR. Mechanistic events that underlie tropism switching can then be reconstructed from the resulting 'haplotype' profiles. We observed that information is not transmitted evenly across VR. Figure 2a shows the patterns of transmission accompanying BPP → BMP/BIP or BMP → BPP tropism switching. In both cases, 3' markers were transmitted with 100% efficiency whereas 5' markers were transmitted at frequencies approaching 50%. Variability at adenines correlated with the transfer of proximal substitutions, whereas lack of variability correlated with their absence. In several cases, we observed mosaic patterns in which stretches of variable, TR-derived sequence were interrupted by non-variant, VR-derived sequence (filled circles, Fig. 2a). Together, these results argue against a simple

cut-and-paste mechanism as commonly observed in transposition reactions^{10,11}.

Because the sequence determinants that govern receptor specificity are unclear, tropism switching assays are inherently biased by a powerful yet poorly defined set of selective pressures. We therefore recorded substitution patterns using PCR-based *in vitro* assays that select for variability at single, precisely defined positions, with no selection for tropism switching or phage infectivity. These assays are based on the loss of restriction sites in parental VR sequences that result from the transmission of synonymous substitutions in TR. As shown in Fig. 2b, *in vitro* variability assays revealed patterns of marker transfer that are surprisingly different from those we had previously observed. *AflIII*-selected clones preferentially transferred the middle portion of TR containing the selected site (position 37), and transfer frequencies precipitously fell in either direction. *MboII*-selected clones transferred the 3' end of TR, which contains the

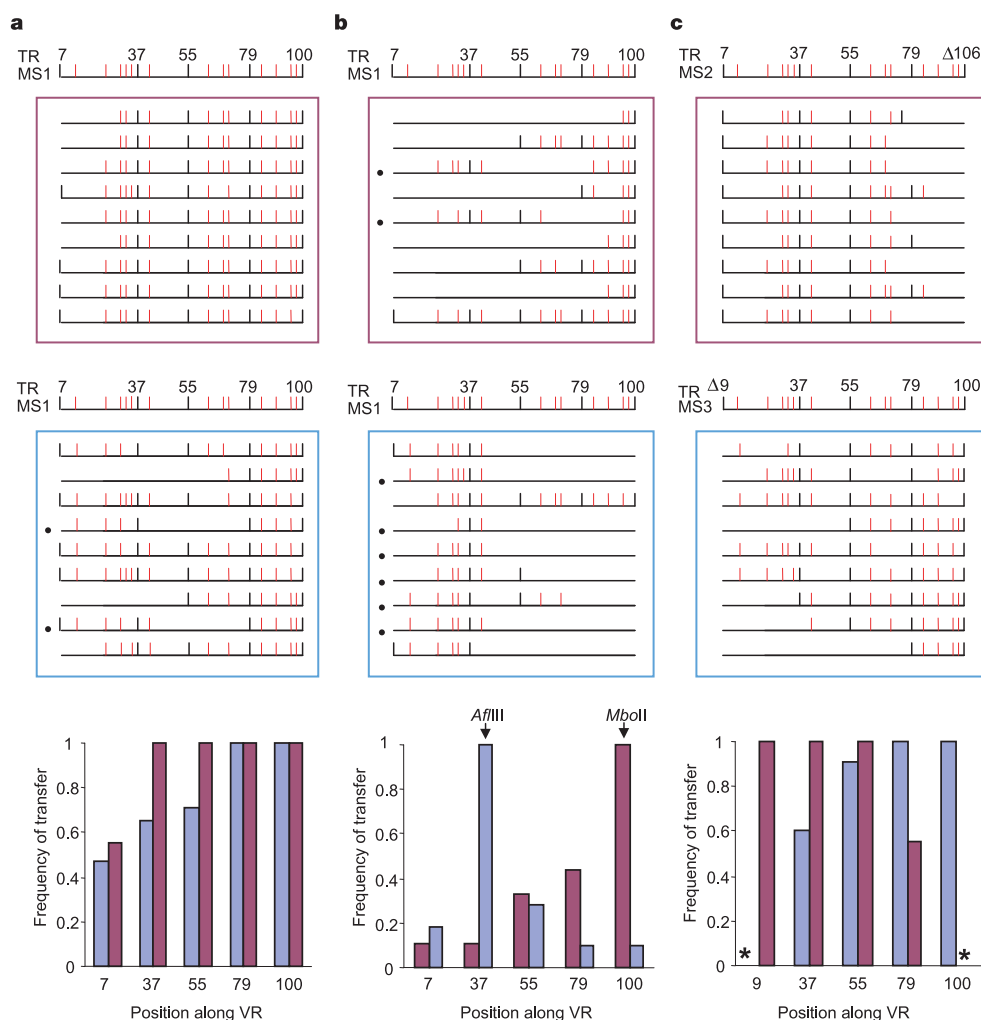


Figure 2 Multiple substitution experiments. **a**, TR of phage MS1 contains synonymous substitutions marked with black lines (see Methods); TR adenines are marked with red lines with adjacent sites represented by a single line. Data boxed in purple or blue schematically represent the VR sequences of nine independent tropism variants. Purple box, BPP-MS1 → BMP/BIP; blue box, BMP-MS1 → BPP. Within the boxes, a black line indicates that a substitution was acquired from TR; a red line indicates that a position varied with respect to the parental VR by adenine mutagenesis. The frequencies of transfer of synonymous substitutions (transmission histograms) are shown at the bottom. Purple bars, BPP-MS1 → BMP/BIP; blue bars, BMP-MS1 → BPP. **b**, *In vitro* variability assays (see Methods) after selection for transfer of synonymous substitutions from TR to VR that confer resistance to *MboII* (position 100, boxed in purple) or *AflIII* (position 37,

boxed in blue). Transmission histograms corresponding to the *MboII* selection (purple bars) or *AflIII* selection (blue bars) are shown at the bottom, along with positions of restriction enzyme cleavage (arrows). **c**, TR of phage MS2 contains a 1-bp deletion at position 106 which, if transferred to VR, results in a frameshift mutation in *mtlD* and non-infectious phage (see Methods). The data boxed in purple depict VR sequences of BPP-MS2 → BMP/BIP tropism variants. TR of phage MS3 contains a 1-bp deletion at position 9, which, if transferred to VR, results in a non-infectious phage. The data boxed in blue show BMP-MS3 → BPP tropism variants. Transmission histograms correspond to BPP-MS2 → BMP/BIP (purple bars) or BMP-MS3 → BPP (blue bars) tropism switching. Asterisks indicate the lack of transfer of frameshift mutations that are subject to negative selection.

selected site (position 100), but were indifferent to sequence variation at the 5' end. In both cases, maximal frequencies of marker transfer were shifted to the exact point of selection. Most events displayed either interrupted patterns of transmission or patches of transmission flanked by invariant sequence (filled circles, Fig. 2b). Despite the lack of selection for mutagenesis, all of the VR sequences in Fig. 2b contain adenine substitutions. This indicates that mutagenesis occurs at a high frequency and homing is rate limiting in the diversity-generating process. To probe further the extent of plasticity, we imposed a strong negative selection against transfer of the 3' or 5' boundaries of TR. This was accomplished by the introduction of frameshift mutations that, if transferred, produce non-viable phage. Notably, the system accommodated these rather extreme selections. Both mutant phages were able to switch tropism while avoiding the transmission of frameshift mutations, generating transmission histograms that are essentially mirror images (Fig. 2c; see also Supplementary Fig. 5).

The observation of flexible mosaicism in response to divergent

selective pressures has direct mechanistic implications. It suggests that a pool of phage variants, most of which contain relatively small random portions of novel sequence, are produced and subjected to selection. Selection at a single position, as imposed by *in vitro* restriction-enzyme-based assays, tends to isolate shorter variable sequences centred around the point of selection. More complex selections for novel receptor specificity select for larger segments of transferred, mutagenized sequence (Fig. 3a). These conditions could be satisfied by a mechanism in which site-specific homing, initiated at IMH, is followed by random gene conversion due to recombination or repair. According to this model, a heteroduplex is formed at VR during the variability generating process (Fig. 3b). The heteroduplex would be characterized by a high density of mismatched base pairs resulting from the hybridization of VR with a TR-derived cDNA. Mismatch repair, or an analogous process, would give rise to chimaeric VRs containing 'patches' of sequence variation.

A consequence of the diversity-generating mechanism is that variability is introduced into the *mtd* locus in a highly targeted manner. Diversification exclusively occurs within the boundaries of the VR, it only occurs at positions corresponding to adenine residues in TR, and it can be limited to the subset of bases that are subject to selection. This 'focusing' of variability has the potential to be highly adaptive as it provides a means to respond efficiently to selective pressures while minimizing the accumulation of unnecessary or deleterious substitutions. The repair step may be essential given the high rate of adenine mutagenesis, and it allows optimization of receptor specificity through iterative rounds of selection^{14,15}.

The ability to diversify protein domains involved in ligand-receptor interactions has broad utility. We therefore suspected that elements homologous to the *Bordetella* phage retroelement might be found elsewhere in nature. To identify related sequences, open reading frames (ORFs) of bacterial origin containing conserved RT domains were compiled. A subset clustered phylogenetically with Brt (Fig. 4a). Adjacent sequences were examined and in all cases candidate TR and VR repeats were identified, with VRs located at the 3' end of an ORF. Further annotation revealed an array of cassettes that we now designate as putative diversity-generating retroelements (DGRs). Although RT domains are highly related, and DGRs share an overall conservation of structural features (Fig. 4b), there is little if any sequence similarity between other components of these related cassettes. Perhaps the most notable observation is that, in every case, VR analogues differ from their cognate TRs almost exclusively at positions corresponding to adenines (Supplementary Fig. 6). This observation provides compelling evidence that these cassettes function to generate diversity in a similar manner. Comparison of the 3' ends of cognate VRs and TRs also suggests the presence of sequences analogous to the *Bordetella* phage IMH site (Supplementary Fig. 6).

DGRs are found in the chromosomes of a wide array of bacterial species and they display interesting variations on a common theme (Fig. 4b). *Nostoc* and *Trichodesmium* species contain cassettes in which a single TR is predicted to supply two different VRs with sequence variability. In such cases, the VRs are part of paralogous ORFs and are identical except for bases corresponding to adenines in TR. In addition, several cyanobacterial species contain multiple DGRs that are not homologous and have been independently acquired. Although the *Bordetella* and *Vibrio harveyi* cassettes are present on prophage genomes, there is no evidence of phage association for the remaining sequences. On the basis of the data in Fig. 4, we propose that DGRs have evolved to perform myriad functions in diverse organisms.

Retroelements such as group II introns¹⁶, retrotransposons¹⁷, retroviruses¹⁸ and human long interspersed nuclear elements¹⁹ share two related characteristics. First, they are highly mobile by means of mechanisms that involve reverse-transcribed RNA inter-

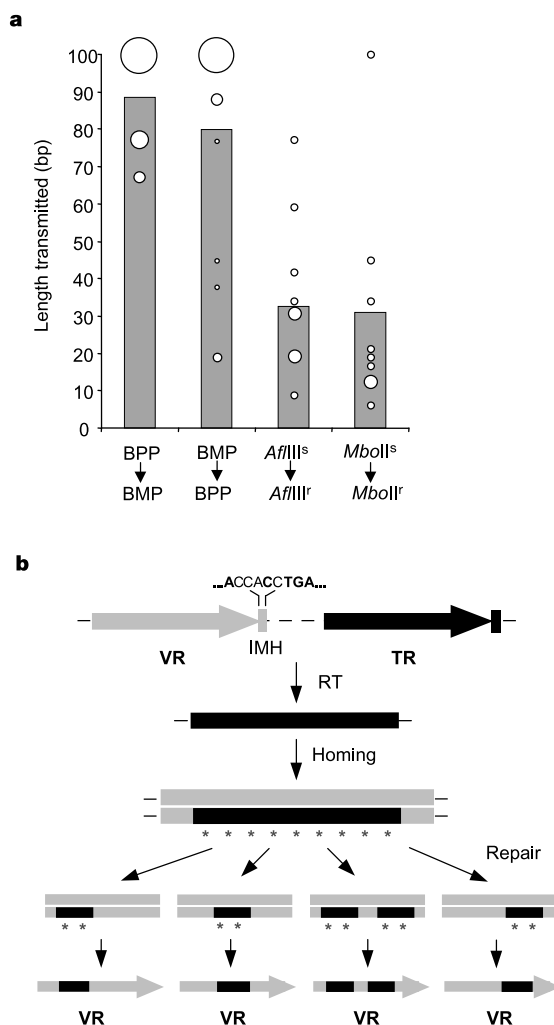


Figure 3 Mosaic VR sequences result from mutagenic homing. **a**, The average length of TR transferred under different selection conditions is shown via a histogram, and the distribution of transferred sequence lengths is depicted with bubbles (area represents the relative number of clones of a given length). Superscript r and s indicate resistant and sensitive, respectively. **b**, In the model, a TR-derived reverse transcript integrates in a homology-dependent manner at VR forming a heteroduplex (asterisks indicate mismatched base pairs), possibly by a mechanism analogous to target-primed reverse transcription^{12,13}. Repair followed by replication would produce mosaic VRs with patches of TR-derived variable sequence.



the equivalent HIV NTD reveals that this similarity is particularly evident for helices 1, 2 and 3 where the root-mean-square deviation (r.m.s.d.) of the backbone atoms is only 1.4 Å (Fig. 1b). Similar superpositions were used to produce the structure-based sequence alignment between MLV, HIV, Rous sarcoma virus (RSV) and human T-cell lymphotropic virus (HTLV) (Fig. 1c). The highest degree of conservation is between residues important in forming the core of each individual $\alpha 1$, $\alpha 2$, $\alpha 3$ three-helix bundle, residues involved in formation and stabilization of the N-terminal β -hairpin and residues that make up $\alpha 6$. In the loops connecting $\alpha 4$ to $\alpha 6$, both the structural and sequence conservation is much less pronounced. Whereas the individual MLV NTD shares a large degree of similarity with the NTDs of HIV, RSV and HTLV, the asymmetric unit of our crystal contains six capsid monomers arranged in a hexagonal ring. Several features of the MLV hexamer in the crystal lead us to believe that this assembly is representative of a protomer of the capsid shell. Given the overall structural similarities between retroviral capsid monomers, the insights into the molecular basis of capsid assembly provided by our structure are relevant not only to MLV but to many other retroviruses including HIV.

The structure of the N-MLV hexameric assembly reveals that the NTDs pack together to form a flat ring 40 Å thick and 90 Å in diameter with a channel that passes through the centre (Fig. 2). The average C α –C α distance across this channel is 20 Å, but where side chains protrude into the cavity this distance is considerably shorter. At its narrowest point, between diametrically related γ -oxygens of serine 18, it is only 12 Å. The overall shape and dimensions of this structure are very similar to those reported for the hexagonal arrangement formed when mature HIV capsid monomers are reconstituted into tubes *in vitro*⁴. Similar hexagonal arrangements of N-terminally His-tagged MLV capsid crystallized in a two-dimensional lattice have also been described⁵. However, because

the presence of the affinity tag and extraneous N-terminal sequences precludes formation of the N-terminal β -hairpin present in a mature capsid^{10–13} and because additional sequences at the N terminus of capsids are known to prevent *in vitro* assembly of tubes^{14,15}, these assemblies are probably more representative of those within an immature virus. Nevertheless, more recent cryoelectron microscopy studies have confirmed the existence of a hexagonal arrangement even in the immature particle¹⁶.

Previous models of HIV hexamer assembly derived from docking of crystal structures into cryoelectron microscopy density⁴ or from deuterium exchange data^{17,18} have suggested that the intersubunit interface is formed by interactions between $\alpha 1$ and $\alpha 2$. In the MLV hexamer structure, neighbouring monomers contact each other through interactions between the structurally conserved helices $\alpha 1$, $\alpha 2$ and $\alpha 3$ (Fig. 3a). The helices pack together to form an 18-helix bundle that makes up the core of the disc. Neighbouring protomers are associated through interactions between residues located on adjacent $\alpha 1$ helices around the underside of the disc and through further interactions between residues located on $\alpha 2$ and $\alpha 3$ on adjacent monomers. Although this difference could represent real variations in the architecture of hexamers from different classes of retrovirus, the degree of conservation at the level of the tertiary structure between the HIV and the MLV proteins suggests that hexagon assembly is probably mediated through structurally equivalent interfaces in these and probably all retroviruses.

In total, the MLV monomer–monomer interface involves 1,100 Å² of buried surface involving residues from $\alpha 1$, $\alpha 2$, $\alpha 3$ and the $\beta 1$ – $\alpha 1$ linker. Apolar buried side chains contribute to the interaction; however, most of the interface is made up from polar contacts through both the main-chain and side-chain atoms. In addition, numerous interactions at the monomer–monomer interface are mediated by water molecules. These solvent-mediated interactions in combination with direct hydrogen bonding create

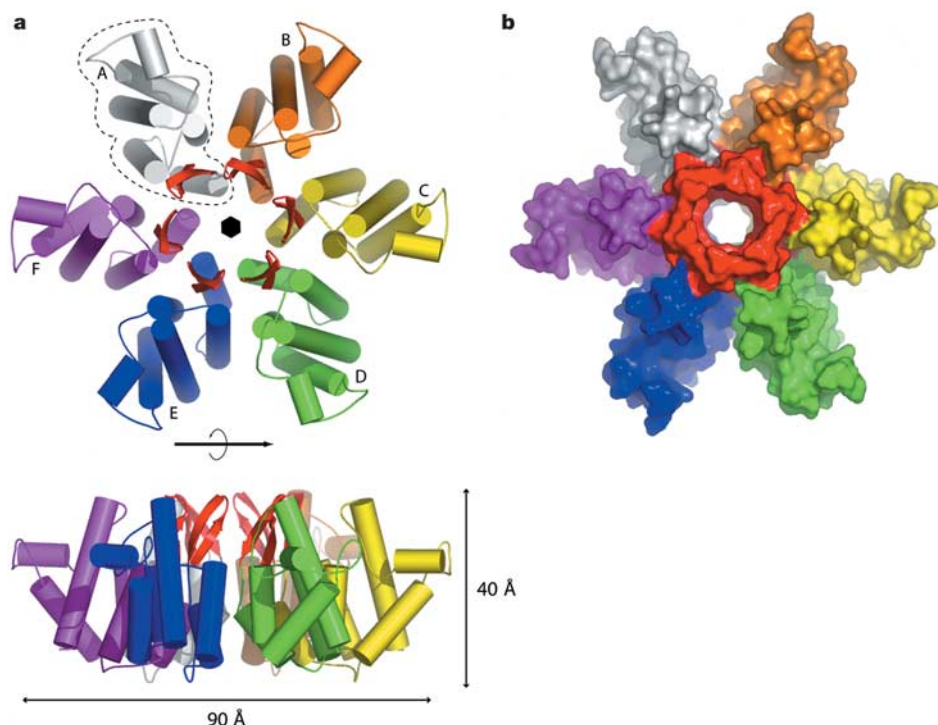


Figure 2 The structure of N-MLV NTD hexameric assembly. **a**, Two orthogonal views of the hexamer. The six α -helices are represented by cylinders and the two β -strands by ribbons. The molecular dimensions of the hexameric assembly are indicated. Individual monomers in the ring are coloured grey, orange, yellow, green, blue and purple and the

$\beta 1$ – $\beta 2$ hairpin is highlighted in red. An individual NTD (grey molecule) is outlined for clarity. **b**, A representation of the molecular surface of the hexamer, individual monomers and the $\beta 1$ – $\beta 2$ hairpin are coloured with the same scheme as **a**.

a network of polar interactions that is pervasive throughout the whole interface (Fig. 3).

With the possible exception of foamy virus, all mature retroviral capsids contain an N-terminal β -hairpin loop, in this structure comprising residues Pro 1 to Trp 13. This loop is only formed upon capsid maturation, brought about by proteolytic processing of the Gag polyprotein precursor. In the structure of an unprocessed form of the HIV Gag¹³, residues 1–11 of the β -hairpin are disordered, demonstrating that this motif does not form in the immature particle. During the maturation process, cleavage releases the free imino terminus of Pro 1, allowing it to form a salt bridge with a highly conserved aspartate, Asp 54, in MLV. In the structure, the Pro 1–Asp 54 interaction is stabilized by hydrogen bonding of the Asp 54 side chain with the hydroxyl of Tyr 12, and by interaction of the carbonyl of Trp 13 with the imino proton of Pro 1. The loop conformation is also stabilized by hydrogen bonding interactions mediated by residues located on the linker between $\alpha 2$ and $\alpha 3$ (Fig. 4a). In contrast, structures of other NTDs reveal considerable disorder around the N-terminal region^{6,8}.

Formation of the $\beta 1$ – $\beta 2$ hairpin loop only occurs at maturation, and mutation of the conserved Asp 54 (and the equivalent Asp 51 in

HIV) prevents mature core formation and markedly reduces viral infectivity^{15,19}. For these reasons, it has been widely speculated that the presence and conformation of this loop are important for capsid assembly, and our structure now clearly reveals why this is the case. The β -hairpins from each monomer are located at the centre of the hexamer and together with residues 14 to 16 that form the linker to $\alpha 1$, collectively form the wall of the central channel. Some side chains from the β -hairpin are orientated so that they protrude into the channel, whereas others are involved in intersubunit contacts as well as the interactions that stabilize both the conformation and position of the loop described above. In this structure, the β -strands are too far separated to form an extended β -barrel but this may be partly due to localized asymmetry imposed by crystal contacts between hexamers packing along the crystallographic 2_1 screw axis. It is possible that these contacts may be more substantial in a perfect P6 hexamer. Regardless, it seems that a major function of the observed arrangement is to stabilize and orient residues involved in intersubunit contacts that are located at the base of the β -hairpin and in the linker. In particular, the side chain C β of Pro 14 packs against the C ϵ of His 48 in the adjacent monomer, and the main chain carbonyl of Phe 15 is hydrogen bonded through a water

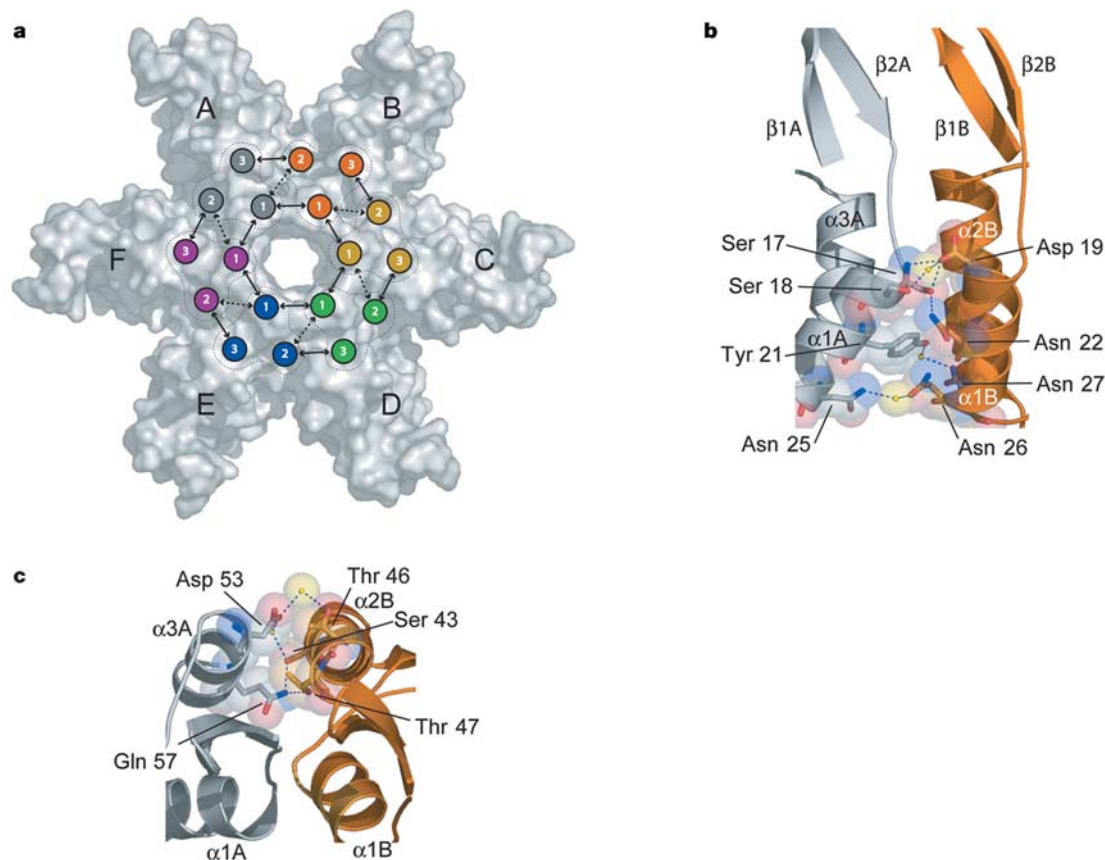


Figure 3 Details of the molecular interactions between monomers at the interface. **a**, A surface representation of the MLV hexamer. The positions of $\alpha 1$, $\alpha 2$ and $\alpha 3$ within each monomer are indicated with the numbered coloured circles following the same colour scheme as in Fig. 2. The solid arrows indicate direct intermolecular interactions between secondary structure elements; the broken arrows indicate indirect water-mediated interactions between secondary structure elements. **b**, Interactions between residues located on neighbouring $\alpha 1$ helices. Residues involved in intermolecular interactions are in stick representation and space-filling models superimposed onto the cartoon diagram representing secondary structures. Water molecules are represented with yellow spheres. The γ -oxygen proton of Ser 17 at the N terminus of $\alpha 1$ forms hydrogen bonds with O $\delta 1$ of Asp 19 and the N $\delta 2$ of Asn 22 on turns one and two of the $\alpha 1$ helix of the adjacent monomer. The interaction is stabilized by another hydrogen bond between the backbone

amide of Ser 17 and the O $\delta 1$ of Asp 19 and by a water-mediated interaction between the carbonyl of Ser 18 and the O $\delta 1$ of Asp 19. In addition, intersubunit association is mediated through an apolar contact in which the phenol ring of Tyr 21 is packed against the aliphatic side chain carbon atoms of Asn 26 of the neighbouring monomer. The phenolic oxygen of Tyr 21 is also involved in a water-mediated hydrogen bond with the N $\delta 2$ of Asn 27. A further water-mediated interaction between the N $\delta 2$ of Asn 25 and the O $\delta 1$ of Asn 26 also contributes to the interface. **c**, Interactions between residues located on $\alpha 3$ and $\alpha 2$ of the adjacent monomers. The N $\epsilon 2$ of Gln 57 located on $\alpha 3$ forms hydrogen bonds with the γ -oxygen of Ser 43 and Thr 47 located on the $\alpha 2$ helix of the adjacent monomer. This interaction is stabilized by water-mediated interactions between the O $\delta 1$ of Asp 53 on $\alpha 3$ with the γ -oxygen of Ser 43 and Thr 46 located on $\alpha 2$ of the adjacent monomer.

molecule with both the N ϵ of His 48 and the O δ 2 of Asp 19 of the adjacent molecule (Fig. 4b). Asp 19 is located at the N terminus of α 1 and along with this interaction is involved in the hydrogen-bonding network between neighbouring α 1 helices. Mutations in many of the positions in and around the N terminus of α 1 in HIV results in the production of altered capsid morphologies^{20,21}. We predict that mutation of Asp 19 in MLV would also have similar detrimental morphological and phenotypic effects. Interestingly, a mutational study using Mo-MLV¹⁵ reported an Asp 63 to Ala point mutation that had phenotypic properties consistent with defects in capsid maturation and assembly. In this initial report, Pro 1 had been incorrectly assigned, and in fact this mutation actually corresponds to a substitution of Asp 19 with Ala, giving us confidence that the intersubunit interactions observed within our structure are relevant to the capsid assembly *in vivo*.

Isolated retroviral capsid cores appear to be unstable^{22,23} and rapidly disintegrate soon after fusion of the viral and cellular

membranes²⁴. However, it seems that the precise mechanism of disassembly varies subtly between different classes of retrovirus^{22,23}. As both C-terminal dimerization and the N-terminal hexamerization interactions are likely to be important for viral disassembly, the differences in disassembly mechanisms may arise from variation in the relative and absolute strengths of these interactions. It has not yet been possible to observe hexamerization of the N-terminal domains of any retroviral capsid proteins in solution. Our structure reveals that much of the interaction between capsid monomers within the hexagonal assembly is a consequence of polar and water-mediated interactions. The absence of a large hydrophobic component is suggestive of only a transient interaction and explains the relative instability of the structure. The weak nature of this interaction is probably important in order to allow the timely disassembly of the viral core upon infection of the host cell. This is highlighted by the fact that the relationship between core stability and virion infectivity appears to be optimally balanced, as mutations that increase or decrease the stability of the capsid core result in a loss of infectivity²⁰.

A clue to the mechanism of core disassembly comes from two recent reports. Cryoelectron microscopy and scanning transmission electron microscopy (STEM) analyses estimate the number of Gag molecules in immature HIV particles at $\sim 5,000$, and that after maturation more than 70% of the capsid molecules in mature virions are in fact not associated with the viral core¹⁶. This observation is further supported by deuterium exchange data that reveal that mature virions contain two forms of the capsid protein¹⁸. The particles contain a slow-exchanging form corresponding to capsid proteins making up the viral core, and another form with exchange kinetics similar to those of free capsid in solution. If this is the case then the concentration of free capsid inside the mature virion is in excess of 3 mM (3,500 free capsid monomers per 145 nm particle), similar to that used in our crystallization experiments. It is tempting to suggest that the weak nature of the capsid assembly requires high concentrations of free capsid in order to maintain a metastable core within the virion, and that after fusion of viral and cellular membranes release of free capsid results in core dissolution.

Mutations that stabilize the capsid core also result in the production of viruses with reduced infectivity²⁰. Furthermore, Arg 110, which constitutes the major determinant of sensitivity to cellular restriction factors such as Fv1 (refs 25, 26) and Trim5 α (ref. 27) is located at the N terminus of α 6, placing the residue on the outer face of the core and therefore making it highly accessible upon cellular entry. At present, the mechanism of host-factor restriction on retroviral infectivity remains elusive and may variously involve capsid destabilization, targeting of capsid to the destruction machinery and other pathways. Our observations now suggest an

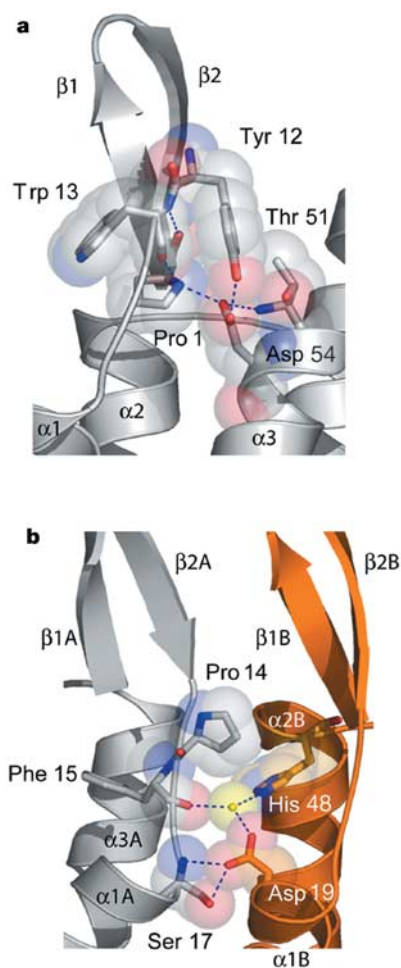


Figure 4 Interactions mediated by residues located on or around the β 1– β 2 hairpin. **a**, Intramolecular interactions that stabilize the position and conformation of the β -hairpin. The imino terminal of Pro 1 is involved in a charge–charge interaction with the carboxylate moiety of Asp 54. This interaction is stabilized by hydrogen bonding of the Asp 54 O δ 1 with the hydroxyl of Tyr 12, and by interaction of the carbonyl of Trp 13 with the imino proton of Pro 1. The loop conformation is also stabilized by a hydrogen-bonding interaction mediated by the amide of Thr 51 located on the linker between α 2 and α 3 and the O δ 2 of Asp 54. **b**, Interactions between neighbouring monomers mediated through residues from the base of the β 1– β 2 loop and the β 1– α 1 linker. The side chain C β of Pro 14 packs against the C ϵ of His 48 in the adjacent monomer, and the main chain carbonyl of Phe 15 is hydrogen bonded through a water molecule with both the N ϵ of His 48 and the O δ 2 of Asp 19 of the adjacent molecule.

Table 1 Structure determination

Data collection	High	Peak	Remote	Inflection
Data set	Se	Se λ 1	Se λ 2	Se λ 3
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Wavelength, λ (Å)	0.9779	0.9779	0.9700	0.9783
Resolution range (Å)	20–1.9	20–2.6	20–2.6	20–2.6
Completeness* (%)	100	99.6	99.9	99.9
R_{sym} † (%)	8.1 (44.2)	4.5 (11.1)	6.0 (16.0)	7.6 (23.7)
MAD phasing				
Resolution‡, d_{max} (Å)	11.2	7.5	6.0	5.2
Mean FoM (20–3.5 Å)	0.72	0.71	0.71	0.68
Refinement				
R_{cryst} § (%)	20.1			
R_{free} (%)	24.1			
r.m.s. bond (Å)	0.010			
r.m.s. angle (deg)	1.2			

* $N_{\text{obs}}/N_{\text{unique}}$ (calculated with merged Bijvoets).

† $R_{\text{sym}} = \sum_i |I_i| / \sum_i I_i$ where I_i is the intensity of the i th reflection and $\langle I \rangle$ is the average intensity.

‡Lowest resolution of data in each bin

§ $R_{\text{cryst}} = \sum_{\text{hkl}} |F_o - F_c| / \sum_{\text{hkl}} |F_o|$

|| R_{free} is as for R_{cryst} but calculated on 5% of the data excluded from the refinement calculation.

additional model whereby natural restriction factors may act by preventing capsid disassembly. In turn, compounds that cross-link capsid components, such as multi-functional reagents designed to target the central channel of the hexamer, might be expected to inhibit retroviral replication. □

Methods

Protein production

The DNA sequence coding for the NTD of N-MLV (capsid residues Pro 1 to Ser 132) was amplified by polymerase chain reaction (PCR) using a plasmid containing the proviral DNA²⁶ as a template. The PCR product was inserted into a pET22b expression vector (Novagen) between the *Nde*I and *Xho*I restriction sites in order to produce a C-terminal hexa-histidine fusion and the sequence Met-Pro at the N terminus. The protein was expressed in the *Escherichia coli* strain BL21 (DE3) and purified using ion exchange, immobilized metal ion affinity and gel filtration chromatography. Verification of the processing of the N-terminal methionine to produce the mature form of the protein was analysed by electrospray ionization mass spectrometry (ESI-MS). The double substitution mutant L4M L126M was prepared using the Quickchange site-directed mutagenesis kit (Stratagene). Seleno-methionine-substituted protein was prepared by expressing the protein in the *E. coli* methionine auxotroph B834 (DE3) grown on seleno-methionine-substituted media.

Crystallization and structure solution

Proteins were crystallized using the vapour batch method. A 16–22 mg ml⁻¹ solution of N-MLV(NTD) or N-MLV(NTD/L4M L126M) in 150 mM NaCl, 20 mM Tris-HCl pH 8.0 was mixed with an equal volume of crystallization solution containing 13–16% PEG 3350, 100 mM sodium citrate pH 5.6. Two-microlitre droplets were dispensed into 96-well vapour batch plates (Douglas instruments) covered with 6 ml of 'Al's oil' (Hampton Research) using an IMPAX 1-5 robot (Douglas instruments). A solution of 10% (v/v) aqueous isopropanol was placed in the side wells of the tray and the drops equilibrated overnight. Next day, the 10% isopropanol solution was replaced with a 20% (v/v) isopropanol solution and crystals typically grew to 0.2 × 0.2 × 0.1 mm³ in one to two weeks. Crystals were collected by transfer into fresh crystallization solution supplemented with 10% (v/v) isopropanol and 15% (v/v) 1,2-propanediol as a cryoprotectant then flash-frozen in liquid nitrogen. The crystals belong to the space group *P*₂₁ (*a* = 86.0 Å, *b* = 78.3 Å, *c* = 85.76 Å, β = 118.9°) with six capsid NTDs in the asymmetric unit. The structure was solved by a three-wavelength MAD experiment using seleno-methionine-containing crystals of the double-methionine-substituted N-MLV(NTD/L4M L126M) on Station 14.2 at the Synchrotron Radiation Source (SRS), Daresbury, UK. All data sets were reduced using the HKL suite of processing software. Twelve selenium atoms were located and the phases refined using the program SOLVE²⁸, resulting in an overall figure of merit (FoM) of 0.61. The local six-fold operators were identified using information from a self-rotation function and the six pairs of Se sites, and were used for averaging and density modification in DM²⁹. The quality of this initial map was good enough to build a polyalanine model into a single monomer using the program O. The hexamer was generated by application of the non-crystallographic symmetry operators to the monomer and the fit improved using real-space rigid-body refinement in the program O. Further refinement against a high-resolution data set collected on the SeMet crystal using REFMAC and ARPP³⁰ resulted in a map into which an essentially complete model was built. The data collection, phasing and refinement statistics are reported in Table 1.

Received 7 July; accepted 5 August 2004; doi:10.1038/nature02915.

- Swanstrom, R. & Wills, J. W. in *Synthesis, Assembly and Processing of Viral Proteins* (eds Coffin, J. M., Hughes, S. H. & Varmus, H. E.) (Cold Spring Harbor Laboratory Press, New York, 1997).
- Briggs, J. A., Wilk, T., Welker, R., Krausslich, H. G. & Fuller, S. D. Structural organization of authentic, mature HIV-1 virions and cores. *EMBO J.* **22**, 1707–1715 (2003).
- Nermut, M. V. & Hockley, D. J. Comparative morphology and structural classification of retroviruses. *Curr. Top. Microbiol. Immunol.* **214**, 1–24 (1996).
- Li, S., Hill, C. P., Sundquist, W. I. & Finch, J. T. Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature* **407**, 409–413 (2000).
- Ganser, B. K., Cheng, A., Sundquist, W. I. & Yeager, M. Three-dimensional structure of the M-MuLV CA protein on a lipid monolayer: a general model for retroviral capsid assembly. *EMBO J.* **22**, 2886–2892 (2003).
- Kingston, R. L. *et al.* Structure and self-association of the Rous sarcoma virus capsid protein. *Struct. Fold. Des.* **8**, 617–628 (2000).
- Cornilescu, C. C., Bouamr, F., Yao, X., Carter, C. & Tjandra, N. Structural analysis of the N-terminal domain of the human T-cell leukemia virus capsid protein. *J. Mol. Biol.* **306**, 783–797 (2001).
- Gamble, T. R. *et al.* Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* **87**, 1285–1294 (1996).
- Gitti, R. K. *et al.* Structure of the amino-terminal core domain of the HIV-1 capsid protein. *Science* **273**, 231–235 (1996).
- Jin, Z., Jin, L., Peterson, D. L. & Lawson, C. L. Model for lentivirus capsid core assembly based on crystal dimers of ELAV p26. *J. Mol. Biol.* **286**, 83–93 (1999).
- Khorasanizadeh, S., Campos-Olivas, R. & Summers, M. F. Solution structure of the capsid protein from the human T-cell leukemia virus type-I. *J. Mol. Biol.* **291**, 491–505 (1999).
- Nandhagopal, N. *et al.* Dimeric Rous sarcoma virus capsid protein structure relevant to immature Gag assembly. *J. Mol. Biol.* **335**, 275–282 (2004).
- Tang, C., Ndassa, Y. & Summers, M. F. Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein. *Nature Struct. Biol.* **9**, 537–543 (2002).
- Gross, I., Hohenberg, H., Huckhagel, C. & Krausslich, H. G. N-Terminal extension of human immunodeficiency virus capsid protein converts the in vitro assembly phenotype from tubular to spherical particles. *J. Virol.* **72**, 4798–4810 (1998).
- von Schwedler, U. K. *et al.* Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. *EMBO J.* **17**, 1555–1568 (1998); corrigendum **19**, 2391 (2000).
- Briggs, J. A. *et al.* The stoichiometry of Gag protein in HIV-1. *Nature Struct. Mol. Biol.* **11**, 672–675 (2004).
- Lanman, J. *et al.* Identification of novel interactions in HIV-1 capsid protein assembly by high-resolution mass spectrometry. *J. Mol. Biol.* **325**, 759–772 (2003).
- Lanman, J. *et al.* Key interactions in HIV-1 maturation identified by hydrogen-deuterium exchange. *Nature Struct. Mol. Biol.* **11**, 676–677 (2004).
- von Schwedler, U. K., Stray, K. M., Garrus, J. E. & Sundquist, W. I. Functional surfaces of the human immunodeficiency virus type 1 capsid protein. *J. Virol.* **77**, 5439–5450 (2003).
- Forshey, B. M., von Schwedler, U., Sundquist, W. I. & Aiken, C. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J. Virol.* **76**, 5667–5677 (2002).
- Ganser-Pornillos, B. K., von Schwedler, U. K., Stray, K. M., Aiken, C. & Sundquist, W. I. Assembly properties of the human immunodeficiency virus type 1 CA protein. *J. Virol.* **78**, 2545–2552 (2004).
- Fassati, A. & Goff, S. P. Characterization of intracellular reverse transcription complexes of Moloney murine leukemia virus. *J. Virol.* **73**, 8919–8925 (1999).
- Fassati, A. & Goff, S. P. Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1. *J. Virol.* **75**, 3626–3635 (2001).
- Karageorgos, L., Li, P. & Burrell, C. Characterization of HIV replication complexes early after cell-to-cell infection. *AIDS Res. Hum. Retroviruses* **9**, 817–823 (1993).
- Best, S., Le Tissier, P., Towers, G. & Stoye, J. P. Positional cloning of the mouse retrovirus restriction gene Fv1. *Nature* **382**, 826–829 (1996).
- Boone, L. R. *et al.* Reversal of Fv-1 host range by *in vitro* restriction endonuclease fragment exchange between molecular clones of N-tropic and B-tropic murine leukemia virus genomes. *J. Virol.* **48**, 110–119 (1983).
- Stremlau, M. *et al.* The cytoplasmic body component TRIM5 α restricts HIV-1 infection in Old World monkeys. *Nature* **427**, 848–853 (2004).
- Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
- Cowan, K. 'dm': an automated procedure for phase improvement by density modification. *Joint CCP4 ESF-EACBM Newsl. Prot. Crystallogr.* **31**, 34–38 (1994).
- Morris, R. J., Perrakis, A. & Lamzin, V. S. ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* **374**, 229–244 (2003).

Acknowledgements We thank S. Gamblin for assistance in crystal handling.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.A.T. (itaylor@nimr.mrc.ac.uk).

corrigendum

Absence of S6K1 protects against age- and diet-induced obesity while enhancing insulin sensitivity

Sung Hee Um, Francesca Frigerio, Mitsuhiro Watanabe, Frédéric Picard, Manel Joaquin, Melanie Sticker, Stefano Fumagalli, Peter R. Allegrini, Sara C. Kozma, Johan Auwerx & George Thomas

Nature **431**, 200–205 (2004).

In Fig. 4e of this Letter, the arrowhead from TSC to Rheb should be a horizontal bar (as from PKB to TSC). In addition, the phosphorylation sites of IRS1 for human should be S312 and S636/639 and for mouse the corresponding phosphorylation sites should be IRS1 S307 and S632/S635. This does not affect any of the results or conclusions of the paper. □

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

US Head Office, New York

345 Park Avenue South, 10th Floor,
New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager/ Corporations:

Peter Bless

Classified Sales Representatives:

Phone: +1 800 989 7718

**Pennsylvania, New York,
Maryland/ Latin America/ NIH:**

Michael McGillion

Midwest USA:

Wade Tucker

East USA/ Canada:

Janine Taormina

San Francisco Office

Classified Sales Representative:

Michaela Bjorkman

West USA/ West Corp. Canada

225 Bush Street, Suite 1453

San Francisco, CA 94104

Tel +1 415 781 3803

Fax +1 415 781 3805

e-mail: m.bjorkman@nature.sf.com

European Head Office, London

The Macmillan Building, 4 Crinan Street,
London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Stefan Hales

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan, Odo Wulffien

e-mail: p.phelan@nature.com

o.wulffien@nature.com

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Harakatemachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: rasami@naturejpn.com

naturejobs

The uncertainty principle

Sociologist Ray Oldenburg believes that everyone needs three places: home, work and a 'third place'. In the United States, the third place is, increasingly, the coffee shop. In Britain, it is without doubt the pub. And in Finland, it's the sauna, as I learned the other month while visiting the country.

Apart from aiding relaxation, saunas play a valuable role in advancing scientific discourse — as I discovered when I tried it out for myself. At a meeting earlier in the day, a reasonably well established scientist had dismissed the idea that the use of short-term grants to support young Finnish scientists is a problem. Everyone gets funded eventually, he said, and the work continues. A younger scientist had begged to differ. Short-term contracts might, at best, distract people from their work, or, at worst, put them off a scientific career in the first place, he argued.

Later that day in the sauna, the younger scientist returned to the argument. He conceded that there was some truth to the older researcher's claim — after all, he had never heard of a graduate student having to leave a programme because funding ran out. But in reality, he said, things like buying a home can become a lot harder when the bank asks you about your job security. And why would you want to train in a country that doesn't guarantee your funding over the course of your education when you can go to one that does?

Uncertainty is a powerful thing. One reason Spain is having difficulty attracting young scientists — even from its own country — is the uncertainty surrounding long-term positions after a researcher completes the much-vaunted Ramón y Cajal programme (see page 488). Spain and Finland may have different 'third places', but they both present their young scientists with a certain amount of uncertainty. If they truly want to attract and keep fresh blood, they will need to dispel this sense of insecurity.

Paul Smaglik
Naturejobs editor



Contents

POSTDOCS & STUDENTS

Spain seeks fresh talent **p488**

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Spain aims at premier league

Could Spain kick off its legacy of inertia and win as much fame for research as for football? It's backing some of its star players with schemes to promote fresh talent, says Quirin Schiermeier.

From Tokyo to Buenos Aires, football fans are thrilled when Real Madrid, Manchester United or Ajax Amsterdam play in Barcelona's Nou Camp. Few of them will know that Europe's largest soccer arena is situated just a stone's throw from Barcelona's university district. And how should they? Unlike the famous FC Barcelona, which hosts a multinational ensemble of football stars, Spanish science suffers both an image problem and a lack of fresh talent from abroad.

Earlier this year, a small group of Ramón y Cajal postdocs gathered to discuss ways in which their programme, Spain's most ambitious effort to deal with this problem, has both succeeded and failed.

The programme — named after the Spanish neurobiologist and joint winner of the 1906 Nobel Prize in Physiology or Medicine, Santiago Ramón y Cajal — was established by the previous Spanish government in 2001. It aimed to address brain-drain problems by attracting young scientific stars, both expatriate Spanish and foreign-born.

Although not unique to Spain, the trend among European scientists to seek work abroad has been especially pronounced there. In the Marie Curie mobility programme run by the European Union (EU), Spaniards rank number one among European researchers looking for a fellowship abroad. From 1999 to 2003, more than 600 Spanish scientists were awarded Marie Curie fellowships, whereas only about 200 young EU researchers went to work in Spain.

In terms of sheer numbers, the Ramón y Cajal programme seems to have succeeded, benefiting nearly 2,000 young scientists. Science in Madrid, Barcelona and other Spanish cities has enjoyed a palpable upswing (see *Nature* **428**, 448–449; 2004).

But numbers alone don't tell the whole story. The programme has a mixed record in attracting foreign scientists and accommodating returners. "Only 17% come from the EU or other countries," admits Carlos Martínez Alonso, president of Spain's national research council, the CSIC. "The programme was meant to attract more foreign scientists." The biggest challenge, all agree, is in finding permanent scientific positions for fellows once their contract expires — the same challenge facing postdocs funded by the EU's Marie Curie programme and national postdoc schemes.

CULTURE SHOCK

Traditionally, Spain has been a difficult place to start a career in science — especially for non-natives. The civil-servant status of most faculty staff makes the system hard for foreigners to penetrate.

"I arrived in Spain last year, knowing little about the country and almost no people," says Giancarlo Franzese, an Italian theoretical physicist, now at the University of Barcelona. "It sounded very appealing, but



Sitting tight: Gumà, Vázquez and Franzese (first, fifth and sixth from left).

then my first grant application was turned down. They said I had no established contacts with the university."

Isolation can be a real problem, Ramón y Cajal researchers have found, particularly if the lines of research they plan to develop have little or no tradition at their host institutes.

Even for Spaniards, applying from abroad for a professorship is next to hopeless, says Ileana Bladé, a climate modeller at the University of Barcelona's astronomy department. The necessary qualification, the *oposición*, is usually awarded to scientists who have been on a 'regular track' in their departments for many years. Most outsiders, even with excellent CVs, come away empty-handed.

Bladé finished her PhD in atmospheric sciences at the University of Washington in Seattle ten years ago, and in 1998 decided to return to her home country. After several short-term projects, she obtained one of the sought-after Ramón y Cajal positions in 2001. She says she would like to stay in research after her five-year contract runs out. But like many other researchers in Spain, she can only hope that the new left-wing government will keep its promise to create more, and more flexible, permanent jobs in science.

"If you have your own research team and funded projects, and you know your way around the Spanish scientific network, Ramón y Cajal may be exactly what you're looking for," says Bladé. "If not, make sure before you apply that you've spotted a lab where people are really interested in you and your experience."

QUIRIN SCHIERMEIER



Painting a rosy picture?
Carlos Martínez Alonso, (left) president of the CSIC (above), says that young researchers are forcing institutes to establish some new priorities.

The government has already announced that the programme will continue beyond 2006. Applicants are invited in all fields — from physics and the life sciences to the social sciences and humanities — provided they have a PhD, an interesting project and an appealing research curriculum.

Although many postdocs around the world meet these requirements, fewer than 20% of Ramón y Cajal positions have been filled by non-Spanish scientists. Yet they would be welcomed. Ramón y Cajal provides cheap and experienced researchers, many of whom have stronger CVs and better publication records than the tenured staff around them. Their salary, some €30,000 (US\$36,000) a year, is paid by the government at first, and then progressively shared by their host institute.

Some institutions show signs of dealing successfully with the programme's — indeed, the country's — two biggest obstacles in building up its scientific workforce: accommodating scientists from abroad, and offering fellows the hope of future employment.

The programme should be measured lab by lab, says Martínez Alonso. The CSIC has a good record of incorporating Ramón y Cajal, he adds, with some 600 Ramón y Cajal researchers currently employed at its 123 research centres. "These young people are really invigorating our institutes, and forcing them to establish new research priorities," Martínez Alonso says. "We do our best to offer them conditions that allow them to make the best of their skills and talent."

At CSIC institutes, such as the National Centre of Biotechnology in Madrid, Ramón y Cajal researchers can either conduct independent research or join an existing group. To make sure they are productive from the very first day, newcomers are offered start-up grants for buying small equipment and receive extensive technical and experimental support.

As soon as they feel competitive enough, they can apply for tenured positions. The CSIC is making good progress with integrating the first generation of Ramón y Cajal researchers: one-third successfully applied for tenure in their first year, in 2002 the number rose to half, and since last year 72% of all candidates are on tenure track.

Their colleagues at universities can only dream of such numbers. One advantage they have over CSIC researchers, though, is that they can acquire teaching experience — which can be helpful when their contracts run out.

If the new government can offer university researchers the same chance as CSIC fellows of tenured employment, and make a clearer path for researchers from abroad, then perhaps Spain will become as famous for its science as for its soccer.

Quirin Schiermeier is *Nature's* German correspondent.

Many Ramón y Cajal scientists feel misled by the suggestion of tenure-track jobs, claiming that the government had given little thought to their long-term career prospects. "At the end of the programme there is no tenure — there's just nothing," says Anna Gumà, a molecular biologist at the University of Barcelona. "We're all waiting for a miracle."

TWEAKING THE RULES

Participants also say that requirements for obtaining grants keep changing. When few foreign scientists applied, for example, the government tweaked the rules so that Spanish researchers who had never spent much time abroad became eligible as well.

"It is all a bit half-baked," says Santiago Vázquez, a pharmacist at the University of Barcelona. "The idea was to keep young researchers in the fridge until the 50-something-year-old Spanish professors retire. But, surprise, most of them are still around, and our generation has the same problem as before."

Despite these criticisms, Ramón y Cajal offers a rare opportunity for young scientists trying to gain a foothold in the rigid Spanish academic system.

Web links

The Ramón y Cajal programme
♦ www.mcyt.es/english_mcyt/CAJAL/default.htm